

Research Article

Performance Analysis of FFBP-LM-ANN Based Hourly GHI Prediction Using Environmental Variables: A Case Study in Chennai

N. B. Sushmi  and D. Subbulekshmi 

School of Electrical Engineering, Vellore Institute of Technology, Chennai 600127, Tamil Nadu, India

Correspondence should be addressed to N. B. Sushmi; sushmi.nb2018@vitstudent.ac.in and D. Subbulekshmi; subbulekshmi.d@vit.ac.in

Received 7 February 2022; Accepted 5 April 2022; Published 23 April 2022

Academic Editor: Araz Darba

Copyright © 2022 N. B. Sushmi and D. Subbulekshmi. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sun is the sustainable and abundantly available alternative resource on the planet Earth. The uncertain nature of the source caused due to various environmental factors increases the need to quantify the irradiation potential at the targeted location, especially for power smoothing processes, solar-fed electrical applications, and the utility grid. The irradiation measuring devices lead to appropriate maintenance and more expenses and are ineffective under dynamically varying irradiation conditions. Hence, a robust ML-based forecasting technique is sufficient to predict GHI under dynamically varying irradiation profiles caused by various environmental factors. Therefore, this study focuses on short-term (hourly) GHI forecasting using the FFBP-LM-ANN approach. In most of the related articles, the selection of independent variables and designing the network is being a challenge and obstacle in attaining fast and efficient prediction. Such a scenario increases the overall computation complexity and understanding. Hence, this study was intended to use a simple process to crucially select 5 environmental factors and to optimally choose the number of neurons for designing the network for better irradiation prediction. Thus, the network is designed with an input layer of 5 selected environmental variables and one hidden layer with 20 neurons, and hourly GHI in the output layer is performed. The approach is trained using the Levenberg-Marquardt BP algorithm in MATLAB toolbox, with the help of a 4-year dataset received for the rooftop panels of VIT University, Chennai, from NREL. The performance of the model during training and testing was validated and analyzed using 9 performance matrices. As a result, FFBP-LM-ANN satisfactorily predicts hourly GHI for the targeted location based on rRMSE of 7.21%, MAE of 0.042, MBE of 0.000492, R of 0.96, MAPE of 44.4%, MRE of 9.5%, and NSE of 94% obtained under testing process. Moreover, the model has performed much better when compared with 9 related models that exist in literature based on the input variables used, network design, epoch, and RMSE. Subsequently, such a predictor will be adaptable and more suitable for the explicit prediction of hourly GHI for different regions across the world having varying climatic conditions, since the study model is designed for locations facing robust climatic nature. More importantly, the designed model is superior with only environmental variables, which are rarely found in the article, rather than geographical variables, which are predominantly used in most of the related literature.

1. Introduction

The drastic increase in greenhouse gases and energy demand has taken a new revolution in energy production over a few decades [1]. As a result, the usage of fossil fuel-based resources is getting depleted day by day. This shoots up the popularity of using renewable energy sources worldwide, and also, this paved the way to tackle the power crisis issues

faced by more developing and developed nations [2]. Among the renewables, solar power generation creates attention among researchers, industrialists, and rural developments due to clean, sustainable, and abundant availability. As the solar generators are purely dependent on irradiation as the main parameter for energy production, the irradiation is uncertain and nonstationary due to highly varying environmental conditions. Hence, highly sensitive measuring

instrument should be essential in PV sites, which are expensive, causes maintenance issues, and are less effective during dynamically varying environmental conditions. But accessing the availability of exact irradiation at the desired location should be accurately known for the power smoothing process, solar-fed electric vehicles/appliances, and also to tackle various grid-related problems [3]. Therefore, the knowledge of knowing short-term irradiation potential is essential to designers for the effective utilization of resources before building a panel for the appropriate applications [4, 5]. There are numerous reviews that have been conducted on various Machine Learning-based irradiation forecasting techniques in the development of solar PV systems [6–16]. It summarizes that ANN predictor finds superior performance under all climatic and environmentally varying conditions when efficient training algorithm, architecture, and specifically selected attributes were incorporated. It also has the dominating features of performing faster computation and handling highly nonlinear datasets compared with other approaches [17, 18]. Therefore, this article was intended to focus on effective training algorithm, simple attributes selection process, and the quicker network design for ANN in attaining the efficient prediction were discussed.

The training algorithms used in the backpropagation process of ANN are Gradient Descent (GD), Bayesian Regularization (BR), Levenberg–Marquardt (LM), and scaled conjugate gradient (SCG) [19]. Some of the neural networks, which undergone analysis with various backpropagation training algorithms, are as follows [20]: conducted FFBP-ANN model to predict daily average GSR using metrological data such as maximum ambient air temperature, minimum ambient air temperature, minimum relative humidity, and day of the year for the site in Madura city, Tamil Nādu. Here, 3 ANN models were designed with different combinations of features mentioned above using two backpropagation training algorithms such as LM and GD. Finally, it was concluded that the FFBP-LM model with minimum air temperature and day of the year outperforms the other models based on MAPE and MSE obtained. Reference [21] proposed two FFBP-ANN with different combinations of the dataset for prediction monthly average GSR using 9 metrological variables such as latitude, longitude, altitude, year, month, mean ambient air temperature, mean station level pressure, mean wind speed, and mean relative humidity for 5 different locations in India (Bangalore, Chennai, Kolkata, New Delhi, and Mumbai). The models were designed with different BP training algorithms such as GD, LM, RP, and SCG. The optimal architecture was designed based on minimum MSE and higher R-value attained by checking the performance with a different number of neurons in the hidden layer. It is concluded that both ANN models outperform well using the LM training process with better MAE, RMSE, and R values. In [22], the average daily solar radiation prediction in Kuwait City was performed by the FFBP-ANN model. The dataset for 5 years was received from 5 different sites in Kuwait. For better generalization and to avoid overfitting issues, the ANN model with GD and LM algorithm for 10 neurons in the

hidden layer is developed and gets MAPE = 86.3 and MAPE = 85.6, respectively. Further, the model with neurons of 1460 has been designed and obtained a MAPE of 94.75. Hence, it is finalized that better performance has been achieved in the case of the LM-ANN model. Likewise, the research proved the effectiveness of the Levenberg–Marquardt (LM) based learning process used in backpropagation. Therefore, it has been taken under consideration by more researchers and industrialists in using LM based backpropagation algorithm for ANN in the irradiation forecasting fields. Some of its contributions are as follows [23]: it performed FFBP-LM-ANN with multiple hidden layers for monthly average GSR prediction in Nigeria using 5 metrological variables such as min and max temperature, mean relative humidity, wind speed, and sunshine hours. The performance was evaluated using statistical tools such as RMSE, R^2 , and MAPE. Reference [24] conducted FFBP-LM-ANN-based GHI prediction for new Delhi using Latitude, Longitude, Elevation, and meteorological parameters such as Months of a year, Days of a month, Temperature, Atmospheric Pressure, Humidity, and Wind Speed. The performance was evaluated using MSE and R . [25] proposed a hybrid Boruta Algorithm and FFBP-NN for hourly global radiation prediction using 13 features such as the month of the year, day of the month, an hour of the 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 of Buraydah in Saudi Arabia. Here, the feature selection process is performed by dividing the entire dataset into three sets of input features by which the optimal input features set is retrieved by the Boruta Algorithm based on the prediction accuracy obtained by a neural network. The performance is evaluated using MAPE, MSE, RMSE, and R^2 . Reference [26] performed FFBP-LM-ANN method with multiple hidden layers for hourly GSR prediction using 5 (selected variable using Cosine Amplitude Method (CAM)) metrological variables such as hourly actual pressure, wind speed, wind direction, relative humidity, and average temperature. To reduce the variance and increase the prediction accuracy, the optimal parameters of the network were chosen based on the trial-and-error method. It was performed by designing 20 ANN models with different combinations of parameters and functions, whose performance is evaluated using correlation coefficient and MSE. Reference [27] conducted spatial, temporal, annual, and 2-year (2018 and 2019) ahead solar radiation prediction using FFBP-LM-ANN approach by considering input variables such as latitude, longitude, day of the year, and year collected for 36 data points in Nigeria from 1979 to 2014. The optimal design of the model was chosen based on the RMSE value attained during testing. The performance of the model was evaluated using R^2 . In [28], the daily solar radiation prediction using the FFBP-LM-ANN model was performed for the site in Hamirpur, Himachal Pradesh. The optimal design with minimum error and high R -value is obtained by considering three network models with different combinations of input variables and neurons. Finally, the model with temperature, humidity,

barometric pressure, rainfall, and sunshine hours has outperformed the others with $R = .86$ and $MAPE = 16.45\%$. Reference [29] performed prediction of hourly GSR using FFBP-LM-ANN using metrological parameters for the site in Kathmandu, Nepal. For obtaining the optimal design and performance, five models of ANN were designed based on considering the different combinations of input variables. Hence, the model with average temperature, relative humidity, sunshine duration, and rainfall amount has outperformed the other model based on good $MBE = 0.0368$, $MPE = 0.1243$, $RMSE = 0.2781$, and $R = 0.9880$. In [30], the daily GSR forecasting was performed by FFBP-LM-ANN with multiple hidden layers using metrological variables and particulate matters for the site in Tehran. The 12 ANN model was designed with different combinations of activation functions. The model with max and min daily temperature, relative humidity, wind speed, and particulates of matter with $\text{tansig-logsig-purelin}$ has outperformed with $MAPE = 3.13$, $RMSE = 0.077$, and $R^2 = 0.97$. Reference [31] performed 6 ML-FFBP-LM-Neural Networks with 32 different combinations of input features such as minimum temperature, maximum temperature, the difference in temperature, sunshine hours, theoretical sunshine hours, and extraterrestrial radiation for predicting monthly average GSR for the site in India. It is confirmed that ANN with DT, H_o , S_o and DT, H_o outperforms well in predicting GSR based on $MAPE = 4.19\%$, $MPE = 3.30$, $RRMSE = 4.90$ and $MAPE = 2.61\%$, $MPE = 1.32$, and $RRMSE = 3.96$, respectively, when compared with other ANN models and 4 empirical models.

The works related to FFBP-LM-ANN reveal that every model can predict irradiation with satisfactory error measures. It is to be noted that, for every location, the independent variables considered should not be similar for accurate irradiation prediction. It varies depending on climatology, environmental conditions around the location of the PV site. Therefore, in every article, the feature variables were manually taken or performed several network models with different combinations of input features for obtaining a better model with the minimum error or used complex feature selection algorithms incorporated with the network design. Such scenarios lead to inaccurate prediction, increase the computational complexity in terms of processing time and data storage, several experimentation processes, and increase the overall complexity of the system. Moreover, no mathematical expression was explicitly provided in any article for choosing the number of neurons in the hidden layer [20, 27]. Generally, they use the trial-and-error method with the consideration of producing minimum statistical error measures. Hence, the attributes and number of neurons should be crucially selected without compromising the performance.

The ultimate goal of this study is to develop a data-driven based irradiation forecasting model for accessing the irradiation potential during a dynamic change in irradiation conditions due to environmental factors without the use of any costly instruments. Therefore, the contribution of the study is as follows: (1) extract the hourly metrological data of 4 years (2016–2019) from the National Renewable Energy

Laboratory (NREL) for the targeted location of VIT University, Chennai. (2) Apply Pearson correlation coefficient as the feature selection process, which helps in the better selection of 5 important features among 18 feature variables. (3) Use a mathematical expression to optimally design the number of neurons for the application as per the dataset availability. (4) Focus on the most powerful Feedforward Backpropagation Levenberg-Marquardt ANN-based solar irradiation prediction model for finding hourly GHI under dynamically irradiation conditions due to various environmental parameters for the rooftop panels of VIT University, Chennai. (5) Performance of the prediction model was evaluated by comparing the obtained GHI with the actual GHI for the corresponding dataset using statistical tools.

The article is organized as follows: the first section covers the overview of the technique performed, description of study location, data collection, and preprocessing procedure, attribute selection, and network design used for prediction process in this study, which were discussed elaborately. The second section concentrates on analyzing the performance of the model and comparison with similar existing works that are effectively performed. The final section projected the overview of the study made with added future ideas.

2. Materials and Methods

2.1. Artificial Neural Network. ANN is the data-driven based learning process that got inspired by the information exchange process carried out in the human brain. It is unique and flexible in solving underlying complex nonlinear functions with high accuracy and faster learning. The basic principle behind every NN is artificial neurons. It constitutes the process element, which receives the input signal and generates the output for the neighboring process after performing with activation function and associated weights [32]. Figure 1 illustrates the architecture of a simple artificial neuron.

The neurons receive the input signal called the independent variable (x_1, x_2, \dots, x_n) and receive the output Y . The output is expressed by the following equation [33]:

$$Y = f(v(w, x)), \quad (1)$$

$$v(w, x) = \sum_{k=1}^n w_k x_k + b = W'X + b. \quad (2)$$

“ v ” is the function used for representing the weights associated with each input and the sum of bias, where w_1, w_2, \dots, w_n are the weights associated with each connection. The vector representation of weight is given as ($W' = w_1, w_2, \dots, w_n$), the input vector is given by ($X = x_1, x_2, \dots, x_n$), and “ b ” represents the bias expressed in equation (2). “ f ” represents the activation function, through which the output is generated from the immediate previous proceeded layer, which is chosen based on the presence of node in the hidden or output layer and the type of problem stated. Some of the commonly used activation functions in the literature are logistic sigmoid function, hyperbolic tangent

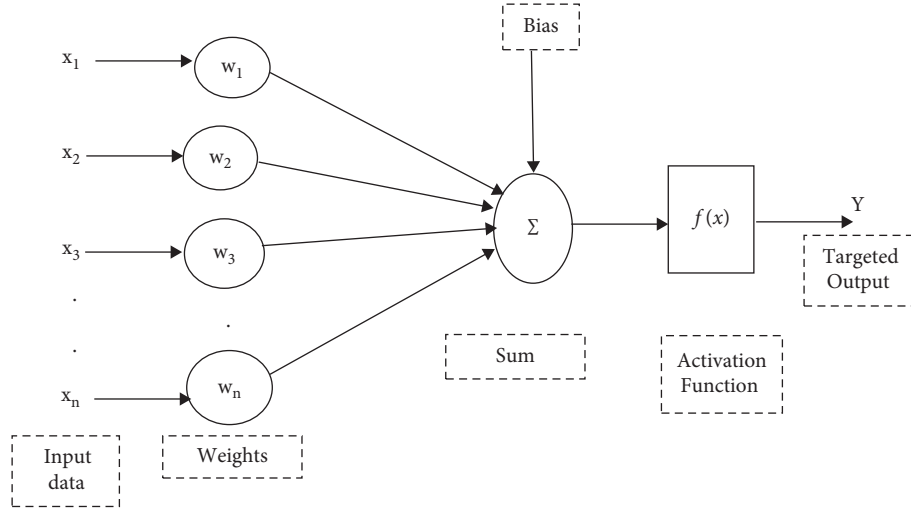


FIGURE 1: Architecture of an artificial neuron.

sigmoid function, Gaussian radial basic function, linear Unipolar step function, Bipolar step function, Unipolar linear function, and Bipolar linear function [34], and their illustration is given in [35].

2.2. Feedforward Backpropagation Process. It is a two-way process, used to train the model for attaining the desired output by reducing the error. The first way includes forward computation of input weights, and the second way includes the backward process of updating weights based on error obtained [36]. The basic structure of the Feedforward network includes an input layer, a hidden layer, an output layer, and their connection associated with its adoptable synaptic weights. Each layer is made up of a specific number of neurons or nodes, by which the information exchange and decision-making take place. Each node receives input from its previous nodes with the weighted sum of their input and added bias. Then, its result proceeded to activation function at each layer for producing the targeted prediction. The procedure is repeated for several iterations with randomly generated weights until the desired output is attained. As it is a black box model [37], the technique finds the useful and complex relationship between the input and output variables in attaining the targeted output. Hence, the learning and generalization have been carried out effectively. Therefore, ANN is especially used in classification, prediction, and especially handling complex and noisy data [38]. Moreover, a neural network with one hidden layer is sufficient and more common than multiple hidden layers-based networks [39, 40].

Later, in advance, to reduce overfitting, convergence at local maxima, and the convergence rate, the back-propagation learning process has been exhibited. This process starts from the output layer and migrates toward the first layer with the updated weights. Here, based on the input signal, the network generates an output. Then, the error is calculated according to the difference between the obtained and actual target. With the note of reducing the error and to get the desired output, the weights are modified and adjusted

using a training algorithm. Hence, the process is repeated on every iteration until the error reaches the predetermined range. The general expression used to calculate the partial derivative of error at the n^{th} layer is given by the following equation:

$$E_n = \frac{1}{2} \sum (Y_n - T_n)^2, \quad (3)$$

where Y_n represents the actual output at the n^{th} layer, T_n denotes the targeted output at the n^{th} layer, and E_n denotes the error generation at the n^{th} layer. Figure 2 illustrates the schematic of the multilayer FFBP-ANN process used in the study with five input features, 20 neurons, and hourly GHI as the output [41].

2.3. Levenberg-Marquardt Training Algorithm. The powerful and widely used algorithm for training the neural network is the LM algorithm for adjusting the weights in accordance with the error calculated. Hence, the error obtained will be minimized within a tolerable range, within less iteration, and in a faster manner. The uniqueness of LM algorithm performance is due to the combined design of the Steepest descent algorithm and the Gauss-Newton algorithm. The computation of the LM algorithm related to the Hessian matrix and the Jacobian matrix is expressed using the following equation:

$$w'_{n+1} = w'_n - [J_n^T J_n + \mu I]^{-1} J_n^T e_n, \quad (4)$$

where

$$H = J_n^T J_n. \quad (5)$$

Based on the parameter μ , the training process switches between the Steepest descent algorithm and the Gauss-Newton algorithm. When μ is small or close to zero, the Gauss-Newton algorithm is performed. It is expressed using the following equation:

$$w'_{n+1} = w'_n - [J_n^T J_n]^{-1} J_n^T e_n. \quad (6)$$

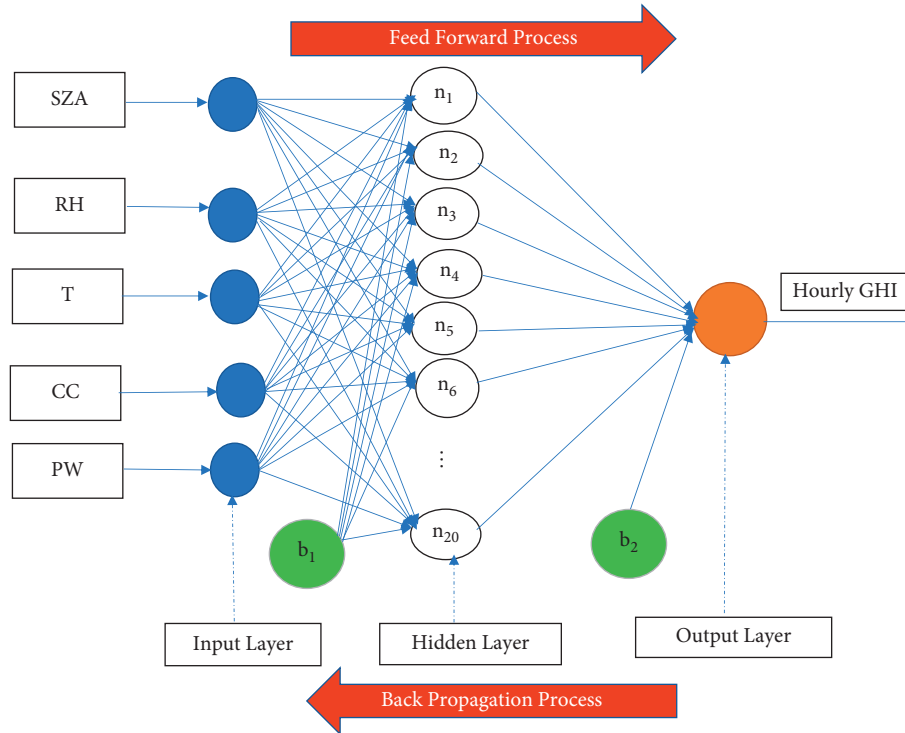


FIGURE 2: Schematic representation of multilayer FFBP-ANN used in the study.

When μ is very larger, the steepest descent algorithm will be carried out. It is expressed using the following equation:

$$w'_{n+1} = w'_n - \alpha g. \quad (7)$$

When μ is still larger, then consider $\alpha = 1/\mu$. In this way, the performance function gets reduced during every iteration, where w'_{n+1} \rightarrow Weight of next iteration, w'_n -Weight of current iteration, J-Jacobian matrix (has the derivative of the network error concerning weight and bias), e_n - Vector of the network error, μ -Combination coefficient, g- Gradient, H-hessian matrix, and I-Identity matrix [34, 42–45].

2.4. Case Study Region. The network model is designed for the Rooftop solar panels of VIT University, Chennai, with a total capacity of 550 KW whose Latitude and Longitude are 12.8406°N, 80.1534°E, respectively, as shown in Figure 3. While looking at the geographical and climatic nature, the region is situated at the southeast coast of India and the northeast corner of Tamil Nadu. It is characterized by tropical wet and dry climate as it is located at the Eastern coastal plane. Moreover, the city lies on the thermal equator, which helps in preventing extremely varying seasonal temperatures. Hence, the City receives a minimum temperature of 18–20 degrees Celsius and a maximum temperature of 38–42 degrees Celsius. Generally, the weather is hot and humid, and the seasons are pleasant (November–February), hot (March–June), and Heavy rainfall (July–September) every year.

2.5. Data Collection. In this study, a 4-year dataset with 18 features, namely, Year, Month, Day, Hour, Temperature,



FIGURE 3: Arrangement of Rooftop solar panels of VIT University, Chennai.

Cloud Type, Dew Point, DHI, DNI, Relative Humidity, Solar Zenith Angle, Surface Albedo, Pressure, Precipitable Water, Wind Direction, Wind Speed, Cell Temperature, and GHI was obtained for the abovementioned location from National Renewable Energy Laboratory (NREL) with the granularity of one hour. The dataset of years 2016, 2017, and 2018 was used for preparing the model. The effectiveness of the design in irradiation prediction is validated by checking the model using the dataset 2019.

2.6. Data Preprocessing. Before processing the data into the model, some of the preprocessing steps, such as removal of outliers and filling the missing data using the most expected GHI values, are performed to improve accuracy and faster response. The outliers were removed from the dataset using the “isoutliers” command to prevent wrong prediction and

skewing of the ML model. Hence, the night data were removed from the entire dataset. Because the power generation from solar panels is performed only during the presence of sunlight, therefore, the data were considered only from 7 am to 5 pm, and others were removed as outliers. The second most important preprocessing step taken is data normalization. It is used to transform the different ranges of the variable between 0 to 1. This process helps improve the accuracy and training process and make a better correlation between features. They are formulated using the following equation:

$$X_{\text{Normalized}} = \frac{x_{\text{Act}} - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)} \quad (8)$$

Once the training and testing using the normalized value are over, then for verification and validation of model performance, the obtained predicted GHI and the target GHI values were generated and denormalized using the following equation:

$$x_{\text{Act}} = X_{\text{Normalized}} [\text{Max}(x) - \text{Min}(x)] + \text{Min}(x), \quad (9)$$

where x_{Act} represents the present value of data, $\text{Max}(x)$ and $\text{Min}(x)$ represent the maximum and minimum value of the dataset, and $X_{\text{Normalized}}$ ranged from 0 to 1 obtained using equation (8).

2.7. Feature Selection Process. As the model is highly dependent on feature variables for reliable prediction, the most significant features for attaining hourly GHI among meteorological data are ranked using the Pearson correlation coefficient. It numerically expresses the relation between the independent variable and the dependent one. The mathematical equation for calculating the coefficient is expressed in the following equation:

$$r = \frac{\sum(x_k - \bar{x})(y_k - \bar{y})}{\sqrt{\sum(x_k - \bar{x})^2 \sum(y_k - \bar{y})^2}} \quad (10)$$

where, r -Correlation coefficient, x_k -Value of the input variable, \bar{x} - mean value of input variable, y_k -values of the output variable, and \bar{y} -mean value of output variable. The 'corrplot' is the MATLAB command used for calculating the Pearson correlation coefficient between each input and output variable. Hence, the computational complexity, cost, and forecasting error get decreased. To avoid overfitting problems in the ML model, only training data are used to find the correlation rank. However, our goal is to predict GHI under dynamically varying irradiation profiles especially due to various environmental factors. Hence, cell temperature, DHI, and DNI ranked in the first 3 places in correlation calculation, whose regression fit is .99902 when considered, as the features for the network are purposefully not considered for this study application. So, we are considering the rest of the superior environmental variables from the Pearson ranking. Therefore, after analyzing the database, the model was designed using selected features such as Solar Zenith angle ($r = -0.88456$), Relative Humidity ($r = -0.65959$), Temperature ($r = 0.55199$), Cloud cover

($r = -0.21625$), and Precipitable water ($r = -0.65959$) for better prediction without sacrificing accuracy. It is to be observed that a higher negative correlation coefficient value also has a more significant relation to GHI prediction. All these preprocessing procedures were carried out in the same MATLAB version mentioned.

2.8. Design of FFBP-LM-ANN Used in the Study. FFBP-LM-ANN-based forecasting mechanism is developed in MATLAB 2020b platform using "nntool" to predict short-term (hourly) GHI under dynamically varying environmental conditions especially using environmental data. Furthermore, for the prediction of hourly GHI, the ANN is designed with 3 layers, such as the input layer, one hidden layer, and the output layer. The number of neurons in the hidden layer whose lower and upper limits are calculated based on equation (11) is as follows [46]:

$$2(n_i + n_o) \leq n_1 \leq \frac{K(n_i + n_o) - n_o}{n_i + n_o + 1}, \quad (11)$$

where n_i indicates the number of input variables, n_o represents the number of output variables, and K represents the number of instances. Usually, for any complex applications, it is enough to have the number of hidden layers that is chosen to be one for a better and simple design [21]. In this work, for the calculation of the number of neurons in the hidden layer, we have considered n_i as 5, n_o as 1, and K as 13198. Therefore, the number of neurons in the hidden layer ranges from 12 to 11312.4, respectively, according to equation (10). For simple and faster access in identifying the number of neurons in the hidden layer, we have considered the neurons that were chosen in the related work with five independent variables. Thereby, the optimal number of hidden layers and number of neurons in the hidden layer are decided during the learning state after performing several trials by considering the criterion MSE, time taken for training, number of Epochs, and best fit R-value. Hence, the ANN model used here is Feedforward Backpropagation Neural Network with five independent variables, a single hidden layer with 20 neurons, and one dependent variable GHI as output. Also, we used Levenberg-Marquardt Backpropagation optimizer as the training algorithm for updating the weights. Even though the memory consumed is more, satisfactory processing at a very faster rate also makes the error within the tolerable range. More significantly, the optimizer stops functioning when there is no improvement in generalization, which are indicated by an increase in the MSE value of the validation samples. The inclusion of the activation function in the network plays a major role in transforming or mapping the complex input to the response variable and decides whether the neuron is to be activated and proceeds as the input to the consecutive layer or not. The most effective activation function used in the hidden layer is "tansig," and "purelin" is used in the output layer [28, 32, 47]. Because "tansig" is the nonlinear activation function used for solving a complex problem in a single layer network, it transfers the real value of input between -1 and $+1$ and supports well under the backpropagation process.

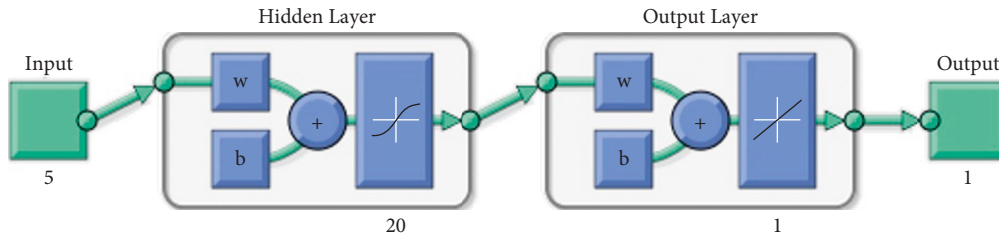


FIGURE 4: The architecture of the designed ANN.

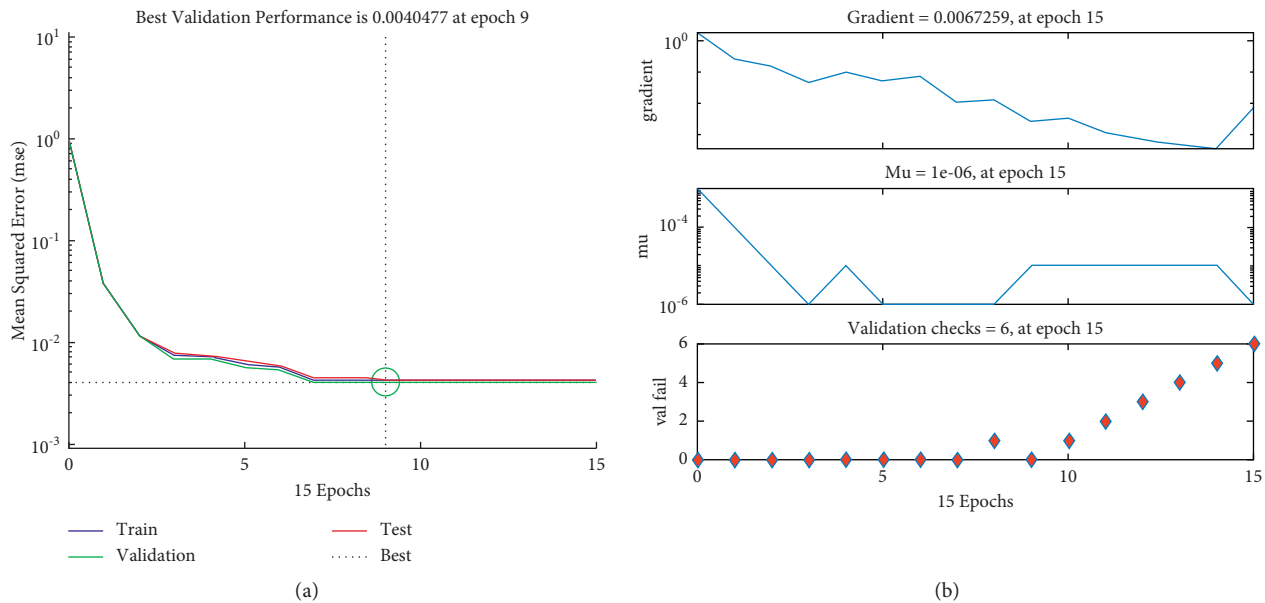


FIGURE 5: (a) MSE plot. (b) Performance of gradient, mu and validation check during the training process of designed FFBP-LM-ANN model for the collected dataset.

Also, it has the capability of strongly mapping the negative input value as negative and also producing zero centered output for the zero-input data compared with the sigmoid function. The “purelin” is the linear activation function that produces the responses between $-\infty$ to $+\infty$. Here, it is preferred in the output layer, since the output is not confined within any range.

The model is processed by considering the dataset of the year 2016, 2017, and 2018 for preparing the model. Among them, it utilizes 75% of data for training, 15% for validation, and 15% for testing. Finally, the 2019 dataset was used for validating the effectiveness of the designed model. In order to observe the performance of the model under all climatic conditions at the targeted location, the data of the entire year are used to train and test the network. Figure 4 illustrates the Architecture of the designed ANN using MATLAB 2020b software.

3. Results and Discussion

3.1. Performance Analysis. The MSE plot of the designed FFBP-LM-ANN model obtained during the training, validation, and testing process related to Epochs is illustrated in Figure 5(a). It shows that once the input is processed, the network produces the predicted output. Then, that output

will be compared with the target with the determination of MSE at the end of each iteration. Once the target and prediction are matched, then the learning process ends, and MSE reaches the minimum value. Thus, the plot reported that as the iteration increases, MSE gradually decreases. Therefore, MSE plays a vital role in checking the performance of the model. It is observed that, after 9 epochs, the system fails six times during training according to the validation check provided. Thereby, the best performance with a minimum Mean Square Error target is achieved at 15 epochs within a second. Figure 5(b) demonstrates the performance of gradient, mu, and validation check steps during the training process of the designed model. The first two plots show the attained values of gradient and mu at every epoch. Hence, the optimal values of gradient and mu are achieved at 15 epochs. The validation plot explains that, after seven epochs, the system fails once, and then the system corrects itself. Further, the system fails six times after nine epochs, thereby the training process stops.

Figure 6 shows the regression plots of the correlation coefficient obtained by the LM-ANN model during training, validation, and testing whose values are $R = 0.97401$, $R = 0.97509$, and $R = 0.97298$, respectively. It shows how closely the target and the output were related. It is observed that the overall regression correlation coefficient obtained is

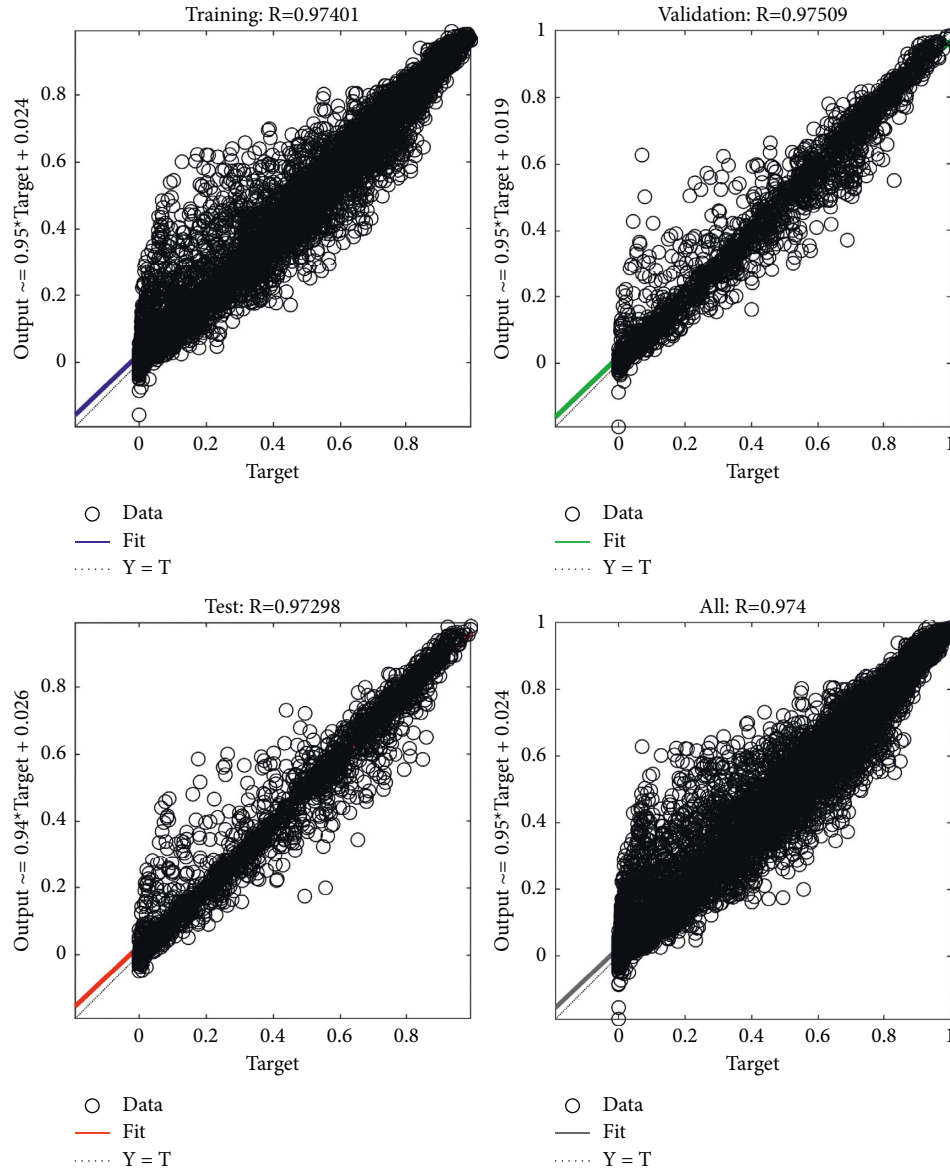


FIGURE 6: Regression plots of FFBP-LM-ANN model during training, testing, and validation process.

$R = 0.974$, which resembles that the output has good relation with the target.

The performance of the designed LM-ANN model was visualized by comparing the actual hourly GHI and the predicted hourly GHI obtained for the targeted location. Hence, Figure 7 shows the actual and predicted hourly GHI during the training process for the VIT site under the corresponding years 2016, 2017, and 2018 (total instances 13198) as illustrated. Then, the difference between the actual and the predicted GHI obtained during the training process is calculated and called an error, as shown in Figure 8. It can be seen that the predicting model can track the actual value firmly, because the notable number of errors is purposely maintained for better generalization during test data and to avoid overfitting problems.

The effectiveness/generalization of the design used in training is evaluated by subjecting the model to a new dataset

with the same 5 features used in training, but for the year 2019, 4410 samples were collected from the exact location where the training data were collected. Figure 9 shows the performance comparison of actual hourly GHI with the predicted hourly GHI produced by the designed model for the corresponding year 2019 for the VIT site are illustrated. The difference between actual and predicted hourly GHI was calculated and graphically visualized as an error in Figure 10. It is observed from the error plot that, in the first half of every year, the prediction is much better; after that, it seems to be more fluctuating. It is due to the monsoon faced by the city every year. Hence, it is confirmed that the design was significantly working well in the hourly GHI prediction for the targeted location of Chennai.

The responses obtained by the FFBP-LM-ANN model during training and testing using selected hourly environmental variables such as Solar Zenith Angle, Relative

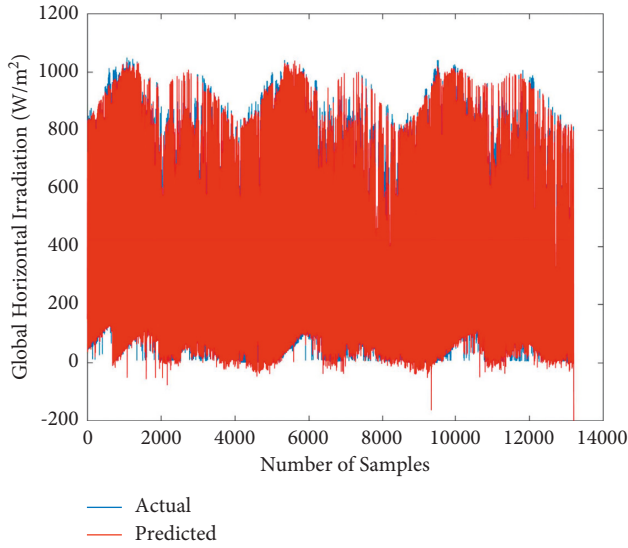


FIGURE 7: Performance comparison of actual and predicted hourly GHI obtained by FFBP-LM-ANN model during the training process.

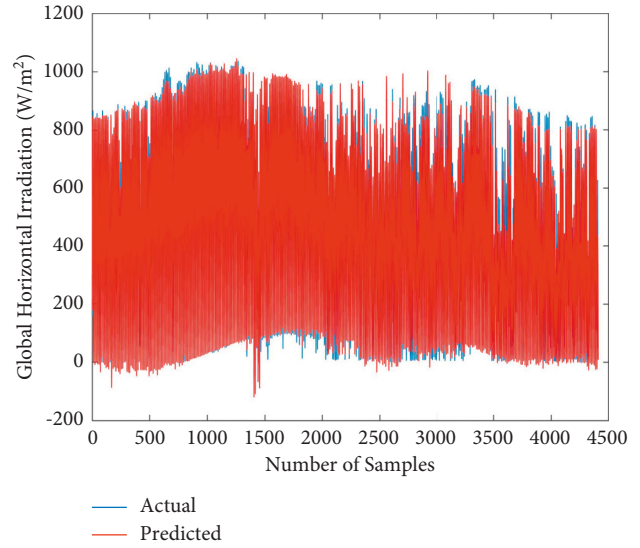


FIGURE 9: performance comparison of actual and predicted hourly GHI by FFBP-LM-ANN model during the testing phase.

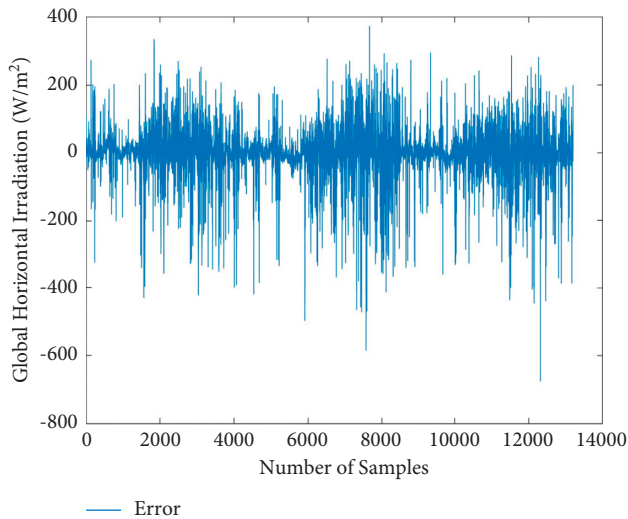


FIGURE 8: Error obtained during the training process of FFBP-LM-ANN.

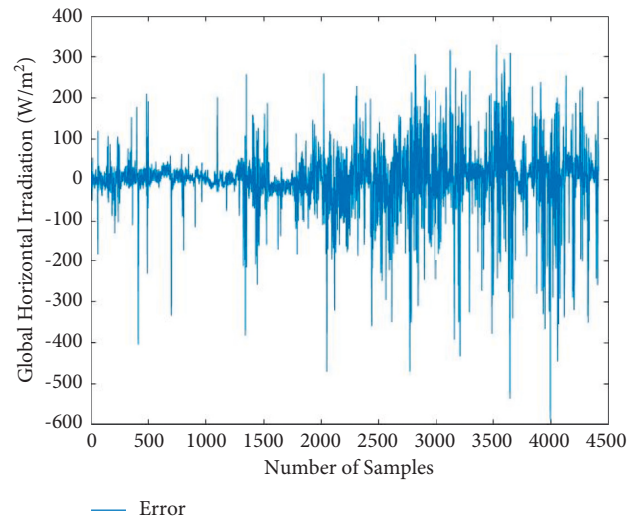


FIGURE 10: Error obtained during the testing process of FFBP-LM-ANN.

Humidity, Ambient Temperature, Cloud cover, and precipitate of VIT University, Chennai, were analyzed. The performances are validated by computing the performance matrices from the actual, predicted, and error values obtained [48–51].

The performance of the study model was evaluated by the most commonly used statistical indicators such as MSE, RMSE, rRMSE, MAE, rMAE, MAPE, MBE, R^2 , MRE, and NSE, which are calculated and displayed in Table 1. According to the range of rRMSE specified in [31], this model attains excellent rRMSE in both training and testing as 6.562% and 7.21%, respectively. Likewise, the ranges of MAPE were specified in [28]. The model receives satisfactory MAPE under training and testing conditions whose values are 0.386397 and 0.44432, respectively. The MBE shows very

small values of bias, in both training and testing; it shows that the model makes a very closer prediction with minor error based on actual and predicted values. The other indicators, such as rMAE and MRE, have got values within the tolerable range. The designed LM-ANN model has attained a correlation coefficient of 0.96 and Nash Sutcliffe efficiency (NSE) of 94% during the testing process. Overall, the model has attained satisfactory performance based on most of the error indicators.

4. Comparison of Study Model with Existing Similar Models

The performance of the study model is compared with other similar existing works performed all over the world mentioned in the literature, which are presented in decreasing

TABLE 1: The statistical performance measure of the FFBP-LM-ANN model during training and testing.

Model/Input features	Performance index	Expression	Training/(2016, 2017, 2018)	Testing/(2019)
FFBP-LM-ANN/SZA, RH, T, CC, Pp	MSE	$1/N \sum_{i=1}^n (G_{ANN} - G_{Actual})^2$	0.004306	0.005078
	RMSE	$\sqrt{1/N \sum_{i=1}^n (G_{ANN} - G_{Actual})^2}$	0.06562	0.07126
	rRMSE	$RMSE \times 100/G_{Actual(Max)}$	6.562%	7.21%
	MAE	$1/N \sum_{i=1}^n G_{ANN} - G_{Actual} $	0.039332	0.042448
	rMAE	$MAE/\overline{G_{Actual}} \times 100$	8.7336%	9.4829%
	MAPE	$1/N \sum_{i=1}^n G_{ANN} - G_{Actual} /G_{Actual} \times 100$	38.6397%	44.432%
	MBE	$1/N \sum_{i=1}^n (G_{ANN} - G_{Actual})$	0.00004002301	0.000492
	R	$\sqrt{1 - \sum_{i=1}^n (G_{Actual} - G_{ANN})^2 / \sum_{i=1}^n (G_{Actual} - \overline{G_{ANN}})^2}$	0.961542	0.959541
	MRE	$1/N \sum_{i=1}^n G_{ANN} - G_{Actual} /G_{total} \times 100$	8.73336%	9.5566%
	NSE	$1 - 1/N \sum_{i=1}^n (G_{Actual} - G_{ANN})^2 / 1/N \sum_{i=1}^n (\overline{G_{Actual}} - G_{ANN})^2$	0.95	0.94

TABLE 2: Performance comparison of study model with similar existing work.

References	Location	Variables	Network	Epochs	RMSE
[25]	Buraydash/Saudi Arabia	DNI, SZA, DHI, hours	4-5-1	-	30.23
[45]	India	Hourly climate data	13-20-1	36	7.591
[27]	Nigeria	Lat, long, day of the year, year	4-20-1	46	3.98
[21]	Mumbai/India	Lat, log and altitude, year and month, mean ambient temperature, mean station level pressure, mean wind speed, mean relative humidity.	6-24-1	1000	3.6461
[23]	Makurdi/Nigeria	Min. Temperature, max. Temperature, relative humidity, wind speed, sunshine hours	5-5-5-4-1	1000	0.98831
[17]	Utter Pradesh/India	Lat, long, altitude, max. Temperature, min. Temperature, humidity, wind speed, sunshine, solar radiation	9-10-1	33	0.1936
[26]	Adane/Turkey	Actual pressure, wind speed, wind direction, relative humidity, average temperature	5-10-30-1	124	0.1543
[24]	New Delhi/India	Lat, long, elevation, the month of the year, day of the month, temperature, atmospheric pressure, humidity, wind speed	9-10-1	1000	0.09441
[30]	Tehran	Temperature, relative humidity, wind speed, particulate matter	5-12-24-1	-	0.0773
Present study	Chennai/India	SZA, RH, T, CC, PW	5-20-1	15	0.07126

order in Table 2. Their performances are compared in terms of the most commonly used matrices called RMSE, epoch, network design, and the input variables used. It is noted that the study model has the exact fit of predicted value to actual data points with a minimum RMSE of 0.07126 than other related works. In addition to that, the model performs the prediction within 15 epochs, which resembles the simple network structure, and an optimal number of neurons were considered, which highlights the superiority of the models compared with other FFBP-LM-ANN designs. Ultimately, the effective environmental feature related to dynamic irradiation profiles considered in this study makes the prediction process more precise when compared with other similar works. More importantly, it is observed that the irradiation can be predicted well under environmental factors rather than geographical features such as latitude, longitude, altitude, elevation, and time zone, which are considered in [17, 21, 24, 27, 45] whose RMSE values and computational complexity are predominantly higher.

5. Conclusion

The hourly Global Horizontal Irradiation prediction was performed by FFBP-LM-ANN using a 4-year database received from NREL for the rooftop panels of VIT University, Chennai. The article discussed a fast and powerful LM-Backpropagation training algorithm for error minimization in the network. The model utilized a simple feature selection process and effective mathematical expression for calculating the optimal hidden layer neurons without compromising the performance. The performance of the model was evaluated using rRMSE, MAE, MBE, R, NSE, MAPE, and MRE of 7.21%, 0.042, 0.000492, 0.96, 94%, 44.43, and 9.5%, respectively. The effective comparison was made with 9 related existing works to show the superiority of the model based on input variables, network design, epoch, and RMSE. Consequently, such a predictor will be reliable and suitable for quantifying the irradiation potential using environmental variables at any location across the world facing dynamic climatic conditions.

In the future, this work will be enhanced by developing the model hybrid with an optimization algorithm for hyperparameter tuning in improving the performance of irradiation forecaster with increased prediction accuracy. In addition to that, the very short-term prediction and seasonal performance of the model should be analyzed.

Data Availability

The data used to support the findings of this study are included in the article. Further data or information required is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

The authors sincerely acknowledge “National Renewable Energy Laboratory (NREL)” for providing a sufficient dataset in the development of our model. Special thanks are due to VIT University, Chennai, for granting facilities and support during the preparation of the manuscript.

References

- [1] S. Manabe, “Role of greenhouse gas in climate change**,” *Tellus A: Dynamic Meteorology and Oceanography*, vol. 71, no. 1, pp. 1–13, 2019.
- [2] V. Manieniyam, M. Thambidurai, and R. Selvakumar, “Study on energy crisis and the future of fossil,” in *Proceedings of the SHEE 2009, Engineering Wing, DDE, Annamalai University, Chidambaram, Tamil Nadu*, pp. 7–12, 2009.
- [3] A. El Hendouzi and A. Bourouhou, “Solar photovoltaic power forecasting,” *Journal of Electrical and Computer Engineering*, vol. 2020, Article ID 8819925, 21 pages, 2020.
- [4] Z. E. Mohamed, “Using the artificial neural networks for prediction and validating solar radiation,” *Journal of the Egyptian Mathematical Society*, vol. 27, no. 1, 2019.
- [5] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, “Solar irradiance forecasting using deep neural networks,” *Procedia Computer Science*, vol. 114, pp. 304–313, 2017.
- [6] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, “A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization,” *Renewable and Sustainable Energy Reviews*, vol. 124, Article ID 109792, 2020.
- [7] P. Singla, M. Duhan, and S. Saroha, “A comprehensive review and analysis of solar forecasting techniques,” *Frontiers in Energy*, 2021.
- [8] S. Ghimire, R. C. Deo, N. J. Downs, and N. Raj, “Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia,” *Journal of Cleaner Production*, vol. 216, pp. 288–310, 2019.
- [9] H. Ghoddusi, G. G. Creamer, and N. Rafizadeh, “Machine learning in energy economics and finance: a review,” *Energy Economics*, vol. 81, pp. 709–727, 2019.
- [10] A. Kumar, M. Rizwan, and U. Nangia, “A hybrid intelligent approach for solar photovoltaic power forecasting: impact of aerosol data,” *Arabian Journal for Science and Engineering*, vol. 45, no. 3, pp. 1715–1732, 2020.
- [11] A. Qazi, H. Fayaz, A. Wadi, R. G. Raj, N. A. Rahim, and W. A. Khan, “The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review,” *Journal of Cleaner Production*, vol. 104, pp. 1–12, 2015.
- [12] G. M. Yaghi, D. Yang, and D. Srinivasan, “Automatic hourly solar forecasting using machine learning models,” *Renewable and Sustainable Energy Reviews*, vol. 105, pp. 487–498, 2019.
- [13] J. Fan, L. Wu, F. Zhang et al., “Empirical and machine learning models for predicting daily global solar radiation from sunshine duration: a review and case study in China,” *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 186–212, 2019.
- [14] A. K. Yadav and S. S. Chandel, “Solar radiation prediction using Artificial Neural Network techniques: a review,” *Renewable and Sustainable Energy Reviews*, vol. 33, pp. 772–781, 2014.
- [15] K. B. Debnath and M. Mourshed, “Forecasting methods in energy planning models,” *Renewable and Sustainable Energy Reviews*, vol. 88, pp. 297–325, 2018.
- [16] A. Alfadda, S. Rahman, and M. Pipattanasomporn, “Solar irradiance forecast using aerosols measurements: a data driven approach,” *Solar Energy*, vol. 170, no. June, pp. 924–939, 2018.
- [17] A. Choudhary, D. Pandey, and S. Bhardwaj, “Artificial neural networks based solar radiation estimation using back-propagation algorithm,” *International Journal of Renewable Energy Resources*, vol. 10, no. 4, pp. 1566–1575, 2020.
- [18] E. A. Ahmed, “Statistical comparison between empirical models and artificial neural network method for global solar radiation at qena, Egypt,” *J. Multidiscip. Eng. Sci. Technol.*, vol. 2, no. 7, pp. 1899–1906, 2015.
- [19] M. Al-Shamisi, A. Assi, and H. Hejase, “Estimation of global solar radiation using artificial neural networks in Abu Dhabi city, United Arab Emirates,” *Journal of Solar Energy Engineering*, vol. 136, no. 2, 2014.
- [20] J. Zhu, B. Arsovska, and K. Kozovska, “Acupuncture treatment in osteoarthritis,” *Int. J. Recent Sci. Res.*, vol. 11, no. 02, pp. 37471–37472, 2020.
- [21] N. Premalatha and A. Valan Arasu, “Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms,” *Journal of Applied Research and Technology*, vol. 14, no. 3, pp. 206–214, 2016.
- [22] M. Bou-Rabee, S. A. Sulaiman, M. S. Saleh, and S. Marafi, “Using artificial neural networks to estimate solar radiation in Kuwait,” *Renewable and Sustainable Energy Reviews*, vol. 72, pp. 434–438, 2017.
- [23] A. Kuhe, V. T. Achirgenda, and M. Agada, “Global solar radiation prediction for Makurdi, Nigeria, using neural networks ensemble,” *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 43, no. 11, pp. 1373–1385, 2021.
- [24] A. Choudhary, D. Pandey, and S. Bhardwaj, “Overview of solar radiation estimation techniques with development of solar radiation model using artificial neural network,” *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 4, pp. 589–593, 2020.
- [25] M. A. A.-H. Abdulatif Aoihan Alresheedi, “Hybrid artificial neural networks with Boruta algorithm for prediction of global solar radiation: case study in Saudi arabia,” *Lecture Notes in Computer Science*, vol. 9, no. 2, pp. 19–27, 2020.
- [26] O. Goncu, T. Koroglu, and N. F. Ozdil, “Estimation of hourly global solar radiation using artificial neural network in Adana province, Turkey,” *Journal of Thermal Engineering*, vol. 7, no. 14, pp. 2017–2030, 2021.

- [27] I. G. Friday, B. C. Udochukwu, T. Igbawua, T. Alaxander, and O. J. Ndubuisi, "Assessment of global solar radiation at selected points in Nigeria using artificial neural network model (ANNM)," *International Journal of Environment and Climate Change*, vol. 9, no. 7, pp. 376–390, 2019.
- [28] S. Kumar and T. Kaur, "Development of ANN based model for solar potential assessment using various meteorological parameters," *Energy Procedia*, vol. 90, pp. 587–592, 2016.
- [29] M. Mohandes, S. Rehman, and T. O. Halawani, "Estimation of global solar radiation using artificial neural networks," *Renewable Energy*, vol. 14, no. 1–4, pp. 179–184, 1998.
- [30] M. Vakili, S.-R. Sabbagh-Yazdi, K. Kalhor, and S. Khosrojerdi, "Using artificial neural networks for prediction of global solar radiation in tehran considering particulate matter air pollution," *Energy Procedia*, vol. 74, pp. 1205–1212, 2015.
- [31] D. V. S. K. Rao K, M. Premalatha, and C. Naveen, "Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: a case study," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 248–258, 2018.
- [32] H. Ghritlahre, "Development of feed-forward back-propagation neural model to predict the energy and exergy analysis of solar air heater," *Trends in Renewable Energy*, vol. 4, no. 2, pp. 213–235, 2018.
- [33] M. O. Moreira, P. P. Balestrassi, A. P. Paiva, P. F. Ribeiro, and B. D. Bonatto, "Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting," *Renewable and Sustainable Energy Reviews*, vol. 135, Article ID 110450, 2021.
- [34] M. A. Jallal, S. Chabaa, S. Chabaa, and A. Zeroual, "Half-hour global solar radiation forecasting based on static and dynamic multivariate neural networks," *Journal of Engineering Research*, vol. 9, no. 2, pp. 203–217, 2021.
- [35] A. P. Yadav and L. Behera, "Solar radiation forecasting using neural networks and wavelet transform," *IFAC Proceedings Volumes*, vol. 47, no. 1, pp. 890–896, 2014.
- [36] N. B. Shaik, S. R. Pedapati, S. A. A. Taqvi, A. R. Othman, and F. A. A. Dzubir, "A feed-forward back propagation neural network approach to predict the life condition of crude oil pipeline," *Processes*, vol. 8, no. 6, p. 661, 2020.
- [37] A. Ibelouad, A. E. L. Kari, H. Ayad, and M. Mjahed, "Multilayer artificial approach for estimating optimal solar PV system power using the MPPT technique," vol. 30, no. December, pp. 109–120, 2021.
- [38] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, 2019.
- [39] H. Pourpasha, P. Farshad, and S. H. Zeinali, "Modeling and optimization the effective parameters of nanofluid heat transfer performance using artificial neural network and genetic algorithm method," *Energy Reports*, vol. 7, pp. 8447–8464, 2021.
- [40] N. Suresh, G. T. Nhinda, V. Hashiyana, P. Veerabhadram, and A. Limbo, "Use of Artificial Neural Network (ANN) to Predict the Output of a Photovoltaic (PV) System," 2020.
- [41] F. A. Makinde, C. T. Ako, O. D. Orodu, and I. U. Asuquo, "Prediction of crude oil viscosity using feed-forward back-propagation neural network (FFBPNN)," *Pet. Coal*, vol. 54, no. 2, pp. 120–131, 2012.
- [42] P. B. Duttagupta and D. Roy, "ANN based techniques for short term load forecasting," *Indian Institute of Technology, Kharagpur 721302*, pp. 327–330, 2002.
- [43] A. Sarfuddin and A. Tiwari, "A comparative study of forecasting solar irradiation of indore city using different machine learning techniques," vol. 6, no. 6, pp. 27–32, 2018.
- [44] B. Plangklang and J. Nantaphunkul, "ANN application for solar radiation forecasting using levenberg-marquardt algorithm," vol. 1, no. 1, 2015.
- [45] A. Choudhary and V. Chauhan, "Solar energy prediction using ANN," vol. 21, no. 1, pp. 1–9, 2014.
- [46] Nastaran Reza Nazar Zadeh, <https://www.udemy.com/course/artificial-neural-network-and-machine-learning-using-matlab/learn/lecture/21837724#overview>, 2021.
- [47] F. J. Diez, L. M. Navas-Gracia, L. Chico-Santamarta, A. Correa-Guimaraes, and A. Martínez-Rodríguez, "ANNs for the daily prediction of horizontal global solar irradiation in an agrometeorological station with mediterranean warm summer climate," *Prime Archives in Agronomy*, Vide Leaf, Hyderabad, India, pp. 1–37, 2020.
- [48] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Engineering Science and Technology, an International Journal*, vol. 21, no. 3, pp. 428–438, 2018.
- [49] A. Takilalte and S. Harrouni, "Daily direct normal irradiance forecasting by support vector regression case study: in ghardaia-Algeria," in *Proceedings of the 2019 Int. Conf. Adv. Electr. Eng. ICAEE 2019*, pp. 1–6, Algiers, Algeria, November 2019.
- [50] Z. U. R. Tahir, M. Azhar, P. Blanc et al., "The evaluation of reanalysis and analysis products of solar radiation for Sindh province, Pakistan," *Renewable Energy*, vol. 145, pp. 347–362, 2020.
- [51] O. Kisi, S. Heddham, and Z. M. Yaseen, "The implementation of univariable scheme-based air temperature for solar radiation prediction: new development of dynamic evolving neural-fuzzy inference system model," *Applied Energy*, vol. 241, pp. 184–195, 2019.