Research Article

Evaluation of Temperature-Based Empirical Models and Machine Learning Techniques to Estimate Daily Global Solar Radiation at Biratnagar Airport, Nepal

Sandeep Dhakal, Yogesh Gautam, and Aayush Bhattarai [iD]

Department of Mechanical and Aerospace Engineering, Pulchowk Campus, Institute of Engineering, Kathmandu 44700, Nepal

Correspondence should be addressed to Aayush Bhattarai; aayush@pcampus.edu.np

Global solar radiation (GSR) is a critical variable for designing photovoltaic cells, solar furnaces, solar collectors, and other passive solar applications. In Nepal, the high initial cost and subsequent maintenance cost required for the instrument to measure GSR have restricted its applicability all over the country. The current study compares six different temperature-based empirical models, artificial neural network (ANN), and other five different machine learning (ML) models for estimating daily GSR utilizing readily available meteorological data at Biratnagar Airport. Amongst the temperature-based models, the model developed by Fan et al. performs better than the rest with an R^2 of 0.7498 and RMSE of 2.0162 $MJm^{-2}d^{-1}$. Feed-forward multilayer perceptron (MLP) is utilized to model daily GSR utilizing extraterrestrial solar radiation, sunshine duration, maximum and minimum ambient temperature, precipitation, and relative humidity as inputs. ANN3 performs better than other ANN models with an R^2 of 0.8446 and RMSE of 1.4595 $MJm^{-2}d^{-1}$. Likewise, stepwise linear regression performs better than other ML models with an R^2 of 0.8870 and RMSE of 1.5143 $MJm^{-2}d^{-1}$. Thus, the model developed by Fan et al. is recommended to estimate daily GSR in the region where only ambient temperature data are available. Similarly, a more robust ANN3 and stepwise linear regression models are recommended to estimate daily GSR in the region where data about sunshine duration, maximum and minimum ambient temperature, precipitation, and relative humidity are available.

1. Introduction

Some of the critical global issues currently encountered by human civilization include global warming and environmental pollution particularly instigated by the excessive use of fossil fuels like petroleum products and natural gas and traditional fuels like timber and firewood [1]. Nepal, being a developing nation where 60% of the entire population is involved in agriculture, has a disproportionate dependence on traditional fuels [2]. However, clean and perpetual solar energy is gaining more and more attention from the government as well as the private sector in recent years. Global solar radiation (GSR) data serve to be one of the critical variables in applications relating to hydrology, meteorology, agriculture, and renewable energy. The GSR is important in the renewable energy sector to predict the capacity and

efficiency of devices based on solar energy applications like photovoltaic cells, solar furnaces, and solar collectors. However, unlike routine meteorological parameters, the data pertaining to daily GSR are not readily available in many locations all over the world [3]. This is particularly relevant for developing countries like Nepal. Lack of measured GSR data has led to the development of several methods to estimate GSR, namely, a neural network [4, 5], empirical models [6, 7], stochastic algorithm [8], and satellite-based methods [9]. Despite the current development in new methods and technologies, the empirical method utilizing meteorological data is preferred because of the cost and technical constraint imposed on new methods and technologies [3, 6, 7, 10].

According to the Department of Hydrology and Meteorology (DHM), only 284 meteorological stations are

currently in operation in Nepal. Out of these, only 64 meteorological stations are equipped with pyranometer to measure daily GSR, while only 34 meteorological stations have the necessary infrastructure to measure daily sunshine duration [11]. The data of daily GSR are not readily available for most of its locations. Thus, various meteorological parameters can be used instead for the estimation of daily GSR.

As of 2019, more than 294 empirical models are available for the estimation of GSR employing readily available meteorological data [7]. Some of the major categories of empirical models include sunshine-based models [12], temperature-based models [13], cloudiness-based models [14], and complex models. Angstrom pioneered the estimation of GSR employing a linear empirical model which was later modified by Prescott [15]. Simplicity of the Angstrom–Prescott (A-P) model and strong correlation of sunshine duration with GSR are the reasons for its extensive application all over the world [12, 16, 17]. After the development of the A-P model, one or more meteorological parameters have been incorporated in the original model to improve the estimation [15]. Sunshine-based models perform more efficiently than models based on other meteorological parameters since the sunshine duration is strongly correlated with the GSR [18–20]. However, the high initial investment and high maintenance cost of the instrument are constraints to the widespread application of sunshine-based models. Therefore, developing empirical models that utilize readily available meteorological parameters such as ambient air temperature, relative humidity, and precipitation are widely preferred.

The ambient temperature range is the most readily available meteorological parameter. One of the simplest temperature-based models consisting of mean monthly maximum and minimum temperature as inputs was proposed by Hargreaves and Samani [21]. After the introduction of the H-S model, several modifications have been developed by incorporating other meteorological parameters to improve the model performance. Hassan et al. modified the H-S model by introducing precipitation term that performed better than two of the most effective sunshine-based models from the literature [13]. Jahani et al. recently developed two new accurate polynomial models that outperformed several temperature-based models from the literature [3]. Although all of the temperature-based models were derived empirically, the variation in ambient temperature was assumed to be largely dependent on the solar radiation arriving on the Earth's surface [22].

Although empirical models are widely analyzed and evaluated, the performance of these models is found to vary according to the geographical location and local climate [6]. Lately, several machine learning (ML) models are employed to estimate GSR at several locations [4, 5]. The capability of generalizing and optimizing time and capacity to resolve problems that are difficult to be represented by an explicit algorithm are some of the biggest advantages of ML models [23, 24]. The main ML models currently in practice include artificial neural network (ANN), support vector machine (SVM), genetic programming (GP), random forests (RF), and adaptive neural-fuzzy inference system (ANFIS). Some of the predominantly applied ANN models include radial basis function network (RBFN) and multilayer perceptron (MLP). Behrang et al. [4] concluded that MLP was more accurate than the RBFN for the estimation of GSR in Iran. Belaid and Mellit [25] applied SVM with different input combinations and concluded that it required a fewer number of input parameters to provide better accuracy than ANN.

The current study presents a comparison between the temperature-based empirical models and ML models. The most common type of feed-forward network, i.e., MLP, is employed in the current study to estimate the GSR at Biratnagar Airport, located in the Eastern Terai Belt of Nepal. Hence, the objectives of the study include the following:

(1) The performance analysis of six different temperature-based empirical models to estimate daily GSR

(2) The application of ANN and other ML models to estimate daily GSR

(3) The comparative analysis of the aforementioned models to recommend the best model for the estimation of GSR

2. Materials and Methods

2.1. Study Location and Data. Nepal is situated between $26.2°$ N and $30.54°$ N latitude in the temperate zone. Nepal experiences 300 days of annual sunshine with an annual average solar radiation of $5 \text{ kWh/m}^2/\text{day}$ [26]. Fourteen-month daily data of various meteorological parameters for Biratnagar Airport ($26.4840°$ N latitude, $87.2662°$ E longitude, and 236 m altitude) were obtained from DHM. The average annual temperature in Biratnagar is $24.3°\text{C}$ with an average annual rainfall of 1898 mm [27].

CMP6 pyranometer is employed to measure the daily GSR on a horizontal surface. The pyranometer consists of a blackened thermopile that absorbs the solar radiation which is converted into heat. Voltage output is generated by the thermopile which is then calibrated to indicate the GSR. A data logger is utilized to record the measured daily GSR. Campbell–Stokes sunshine recorder is employed for the measurement of sunshine duration. Mercury-filled and alcohol-filled meteorological thermometers are used to measure wet-bulb and dry-bulb temperature for normal ambient and low ambient temperature, respectively. Similarly, data for relative humidity is computed using the dry-bulb and wet-bulb temperature.

2.2. Temperature-Based Empirical Models. Several models have correlated ambient air temperature with GSR empirically. Two of the most widely employed empirical model for the estimation of GSR using ambient air temperature data only are the Hargreaves and Samani (HS) model and Bristow and Campbell (BC) model. Thus, the above two models along with other four recently developed highly accurate models are selected in the present study.

Hargreaves and Samani [21] proposed a simple model employing a mean monthly maximum temperature and

mean monthly minimum temperature as inputs for the estimation of daily GSR:

$$\frac{R_s}{R_a} = c_1 (\Delta T)^{0.5}, \tag{1}$$

where $\Delta T = T_{\max} - T_{\min}$ and c_1 is the empirical constant.

Chen and Li [28] developed and analyzed the performance of more than 20 different temperature-based empirical models. Two of the top-performing temperature-based models incorporating ambient temperature range and mean monthly maximum temperature and mean monthly minimum temperature as inputs are taken for the current study. One model incorporates an additional constant term to the original H-S model with an exponent of "1" (abbreviated as Chen and Li (model 1)):

$$\frac{R_s}{R_a} = a + c_1 \Delta T, \tag{2}$$

where a and c_1 are the empirical constants.

Another model is a multiple regression model that takes mean monthly maximum and minimum temperature as inputs (abbreviated as Chen and Li (model 2)):

$$\frac{R_s}{R_a} = a + c_1 T_{\max} + c_2 T_{\min} + c_3 T_{\max} \times T_{\min}, \tag{3}$$

where a, c_1, c_2, and c_3 are the empirical constants.

Bristow and Campbell [29] developed a model considering that the GSR is exponentially related to the ambient temperature range:

$$\frac{R_s}{R_a} = c_1 [1 - \exp (c_2 (\Delta T)^{c_3})], \tag{4}$$

where c_1, c_2, and c_3 are the empirical constants.

Jahani et al. [3] proposed a model considering a polynomial correlation of GSR with ambient temperature range:

$$\frac{R_s}{R_a} = a + c_1 \Delta T + c_2 (\Delta T)^2 + c_3 (\Delta T)^3, \tag{5}$$

where a, c_1, c_2, and c_3 are the empirical constants.

Fan et al. [6] modified the Jahani model by using a different exponent in the ambient temperature range term and incorporating additional average temperature term to improve the model performance:

$$R_s = \left(a + c_1 (\Delta T)^{0.25} + c_2 (\Delta T)^{0.5} + c_3 \Delta T\right) R_a + c_4 T_a, \tag{6}$$

where $T_a = (T_{\max} + T_{\min})/2$, and a, c_1, c_2, c_3, and c_4 are the empirical constants. Extraterrestrial GSR is computed using the following equation [30]:

$$R_a = \frac{24}{\pi} H_{sc} \left(1 + 0.033 \cos \frac{360n}{365}\right) \left(\cos \phi \cos \delta \sin \omega + \omega \frac{\pi}{180} \sin \phi \sin \delta\right). \tag{7}$$

The angle of declination is determined by using the following equation [31]:

$$\delta = 23.45 \sin\left(\frac{360}{365} (284 + n)\right). \tag{8}$$

The day length is calculated as follows:

$$N = \frac{2}{15} \cos^{-1} (-\tan \phi \tan \delta). \tag{9}$$

The sunset hour angle is calculated as follows:

$$\omega = \cos^{-1} (-\tan \phi \tan \delta). \tag{10}$$

2.3. Machine Learning Models. Multilayer perceptron (MLP) model is employed in the current study among several other ANN topologies available in the literature. MLP is particularly useful in modeling to resolve a complex problem. Figure 1 illustrates the structure of MLP that consists of an input layer, a hidden layer, and an output layer. Input signals are multiplied by a set of weights as they are sent to the output layer through the hidden layer.

Typical MLP with a hidden layer can be modeled as follows [32]:

$$y_i = \sum_{j=1}^{6} w_{i,j} x_j + \theta, \tag{11}$$

where θ is the bias in the hidden layer. A nonlinear activation function (typically a sigmoid) is employed for the calculation of output of neurons given by [33]

$$f(x) = \text{sigmoid}(x) = \frac{1}{(1 + \exp(-x))}. \tag{12}$$

Support vector machine (SVM) is a powerful supervised learning technique with excellent generalization ability because of which it is extensively utilized for solving problems regarding pattern recognition, classification, regression, and prediction [34]. SVM function can be used with various kernel functions to implement its regression learner. The application of the Gaussian kernel is popular with SVM classification and regression. Very complex boundaries and relations can be established with the help of Gaussian-based SVM. Medium Gaussian and fine Gaussian are defined based on the slenderness of the Gaussian function being used. Fine Gaussian uses very thin Gaussian function with very low standard deviation. As a result of these thinly separated boundaries, fine Gaussian is susceptible to overfitting. Linear regression learner performs a multivariate linear regression on a set of input data. On the other hand, stepwise linear
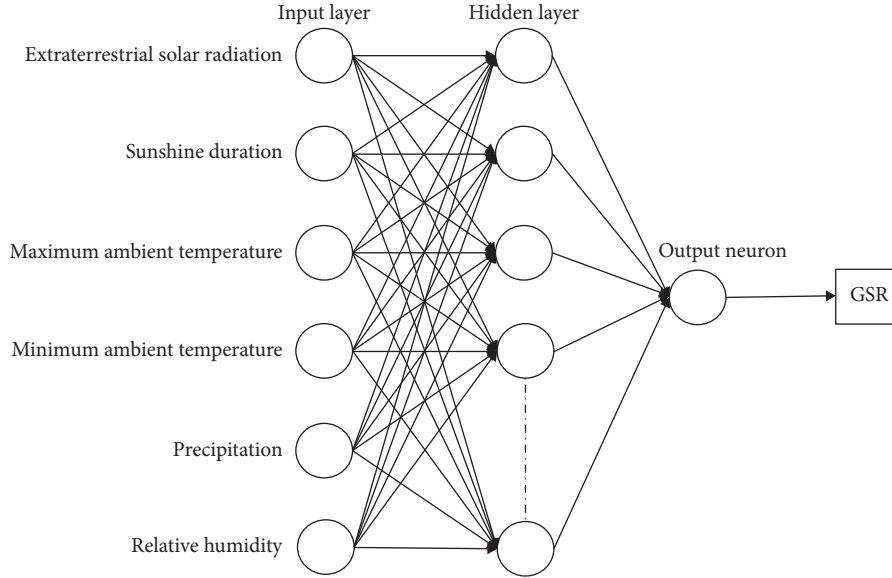
FIGURE 1: The structure of MLP. Six sets of input data are initially fed in an input layer and the desired output is obtained from the output layer through the hidden layer. The number of neurons in the hidden layer is varied, and the corresponding statistical indicators are recorded.

regression utilizes only highly correlated variables for linear regression removing redundant weakly correlated variables.

2.4. Model Training and Testing.

The fourteen-month data is divided into two datasets, initial 85% of the data is utilized for the development of ML models or the calibration of empirical models and the rest 15% is employed for the model assessment. The present study employs extraterrestrial solar radiation (R_a), sunshine duration, maximum ambient temperature ($T_{\max}$), minimum ambient temperature ($T_{\min}$), precipitation, and relative humidity as inputs for the development of ML models.

The performance of the neural network is analyzed by differing the number of neurons in the hidden layer and recording the respective statistical indicators. Neural Net Fitting [35], an implicit application in MATLAB, is employed to design and train the neural network. The most well-known feed-forward network, i.e., MLP, is employed in the current study to model the GSR. Levenberg–Marquardt backpropagation algorithm is utilized to train the network. Training terminates when generalization stops improving, as demonstrated by an increase in the mean square error and the corresponding decrease in R^2. Regression Learner [36], an implicit application in MATLAB, is employed to analyze and evaluate the performance of linear regression, stepwise linear regression, medium Gaussian SVM, matern 5/2 Gaussian process regression (GPR), and exponential GPR.

2.5. Statistical Indicatives.

Four statistical indicators, namely, coefficient of determination (R^2), adjusted R^2, mean square error (MSE), and root mean square error (RMSE) are utilized to evaluate the performance of various models. The performance of the model is evaluated using the following equations:

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} (Y_i - X_i)^2},$$

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^{m} (Y_i - X_i)^2,$$

$$R^2 = \frac{\left[\sum_{i=1}^{m} (X_i - \overline{X})(Y_i - \overline{Y})\right]^2}{\sum_{i=1}^{m} (X_i - \overline{X})^2 \sum_{i=1}^{m} (Y_i - \overline{Y})^2},$$

$$R^2_{\text{adj}} = 1 - \frac{(1 - R^2)(n - 1)}{n - k - 1},$$

(13)

where X_i and Y_i represent the measured and predicted values, while $\overline{X}_i$ and $\overline{Y}_i$ represent the average measured and average estimated values, n is the number of data points, and k is the number of independent regressors.

RMSE value provides an indication of the short-term performance of the model. Lower RMSE value corresponds to better performance. R^2 indicates the variance of the dependent variable that is explained by independent variables.

3. Results and Discussion

3.1. Calibration of Temperature-Based Empirical Models.

Figure 2 illustrates the correlation between the measured value and the estimated value of daily GSR for the calibration dataset. The estimated value of GSR is reasonably correlated with the measured value of GSR for all models. From the statistical indicators (Table 1), the model developed by Fan et al. is better correlated among other temperature-based
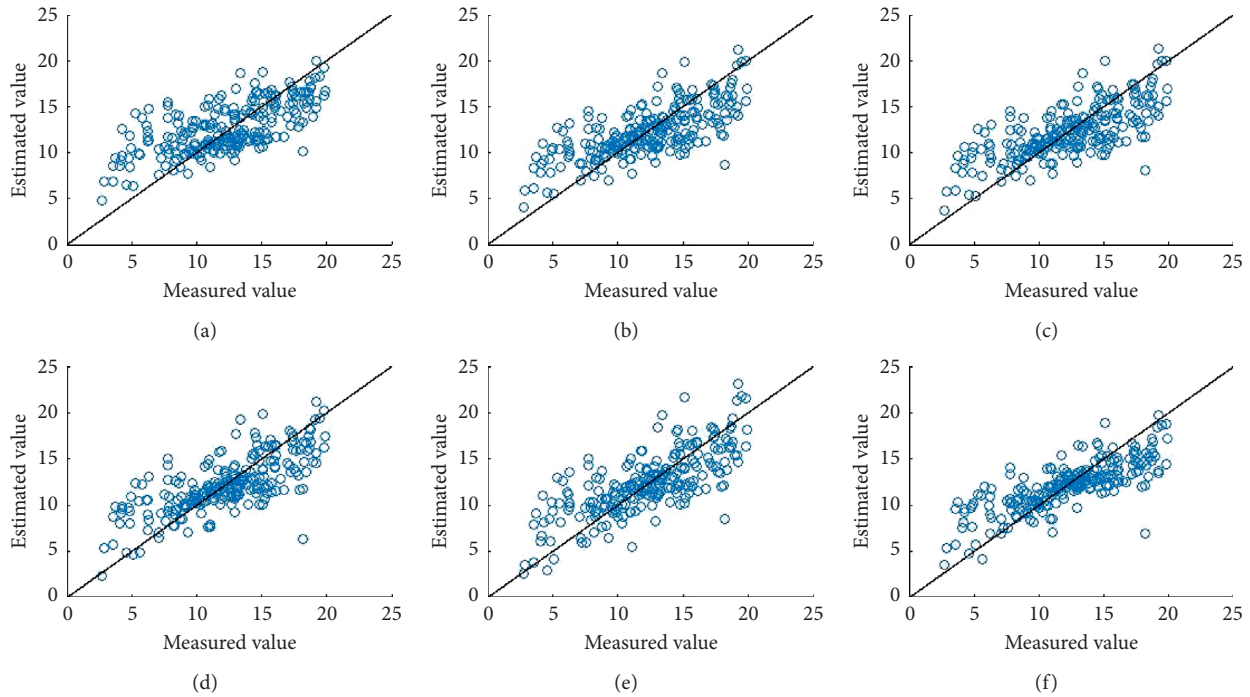
FIGURE 2: Scatter plots of measured daily GSR ($MJm^{-2}d^{-1}$) versus estimated daily GSR ($MJm^{-2}d^{-1}$) using empirical models. The model proposed by Fan et al. shows a strong correlation between measured and estimated value. (a) Hargreaves and Samani (HS). (b) Chen and Li (model 1). (c) Bristow and Campbell (BC). (d) Jahani et al. (e) Chen and Li (model 2). (f) Fan et al.

TABLE 1: Statistical indicators for temperature-based empirical models.

Models	R^2	R^2_{adj}	RMSE ($MJm^{-2}d^{-1}$)
Hargreaves and Samani (HS)	0.5291	0.5291	3.0148
Chen and Li (model 1)	0.5716	0.5697	2.8166
Bristow and Campbell (BC)	0.5764	0.5728	2.7949
Jahani et al.	0.5931	0.5878	2.7455
Chen and Li (model 2)	0.6248	0.6200	2.7336
Fan et al.	0.6435	0.6373	2.5810

models with an R^2 of 0.6435 and corresponding R^2_{adj} of 0.6373. Concerning RMSE, again model developed by Fan et al. performs better than other temperature-based models with an RMSE of 2.5810 $MJm^{-2}d^{-1}$. Chen and Li (model 2) ranks second among temperature-based models with R^2, R^2_{adj}, and RMSE of 0.6248, 0.6200, and 2.7336 $MJm^{-2}d^{-1}$, respectively. The model developed by Fan et al. is reasonable to apply in the region where only data pertaining to ambient temperature range are available.

The least-square method is employed to fit the empirical coefficients for temperature-based models. Empirical coefficients obtained for all temperature-based models are incorporated in Table 2.

3.2. Training and Validation of ANN Models.

Figures 3 and 4 illustrate the correlation between daily GSR estimated by ANN and measured daily GSR for 6 and 10 neurons in the hidden layer, respectively. Table 3 provides a summary of the statistical indicators for ANN models. For the training set, the model with 10 neurons in the hidden layer (abbreviated as ANN5) performs better than other ANN models with an R^2 of 0.8485 and RMSE of 1.4967 $MJm^{-2}d^{-1}$. Likewise, ANN3 ranks second among other ANN models in the training set with an R^2 of 0.8341 and RMSE of 1.6823 $MJm^{-2}d^{-1}$. A comparison of statistical indicators shows that R^2 and RMSE sometimes follow a different trend. Similar to the training set, R^2 and RMSE sometimes follow a different trend in the validation set. Although ANN5 is the best model in terms of R^2, it exhibits a comparatively higher RMSE of 1.8221 $MJm^{-2}d^{-1}$. In the model development, ANN models account for greater variance in the training data in comparison with the temperature-based empirical model.

3.3. Training and Validation of Other ML Models.

Figure 5 illustrates the correlation between daily GSR estimated by various ML models and measured daily GSR. In order to prevent the overfitting of the model, 5-fold cross-validation is performed during the model development. The statistical indicators for these models are incorporated in Table 4.

Medium Gaussian SVM performs better than other models with an R^2 of 0.79 and RMSE of 1.867 $MJm^{-2}d^{-1}$. Similarly, GPR matern 5/2 ranks second with an R^2 of 0.79 and RMSE of 1.870 $MJm^{-2}d^{-1}$. The performance of GPR exponential degrades extensively with cross-validation which might be attributed to the overfitting of the dataset.

TABLE 2: Calibrated empirical coefficients for temperature-based models.

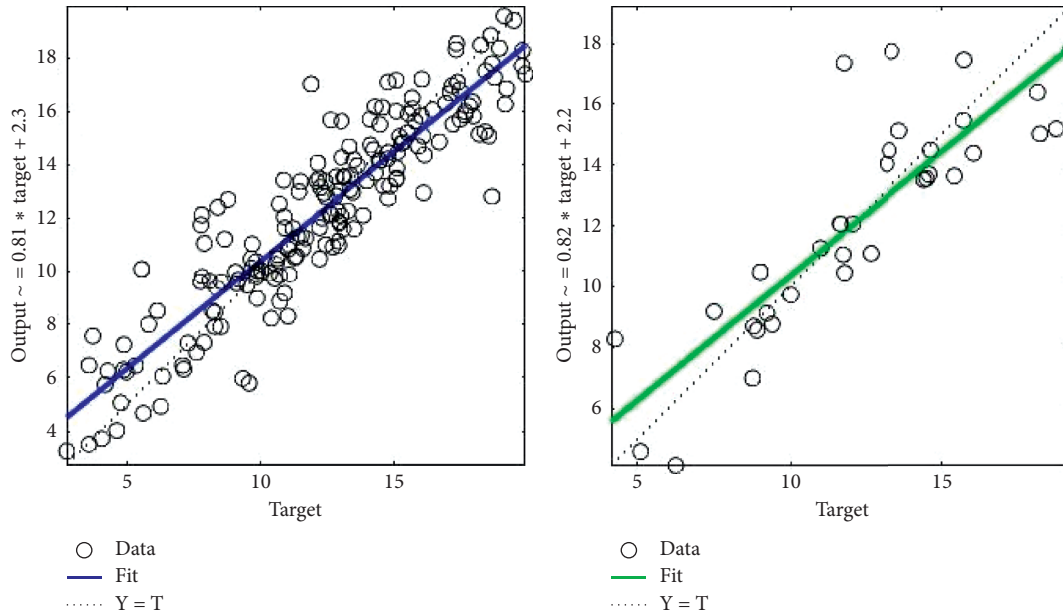| Models	a	c_1	c_2	c_3	c_4
Hargreaves and Samani (HS)	—	0.1274	—	—	—
Chen and Li (model 1)	0.1048	0.0281	—	—	—
Bristow and Campbell (BC)	—	$-9.24E-4$	4.5390	0.1241	—
Jahani et al.	-0.0943	0.0741	-0.0029	$4.945E-5$	—
Chen and Li (model 2)	-0.4274	0.0481	-0.0116	$-6.398E-4$	—
Fan et al.	-2.0060	2.5160	-0.9118	0.0577	0.2966

FIGURE 3: Correlations between GSR estimated $(MJm^{-2}d^{-1})$ by ANN and measured GSR $(MJm^{-2}d^{-1})$. The number of hidden neurons = 6.
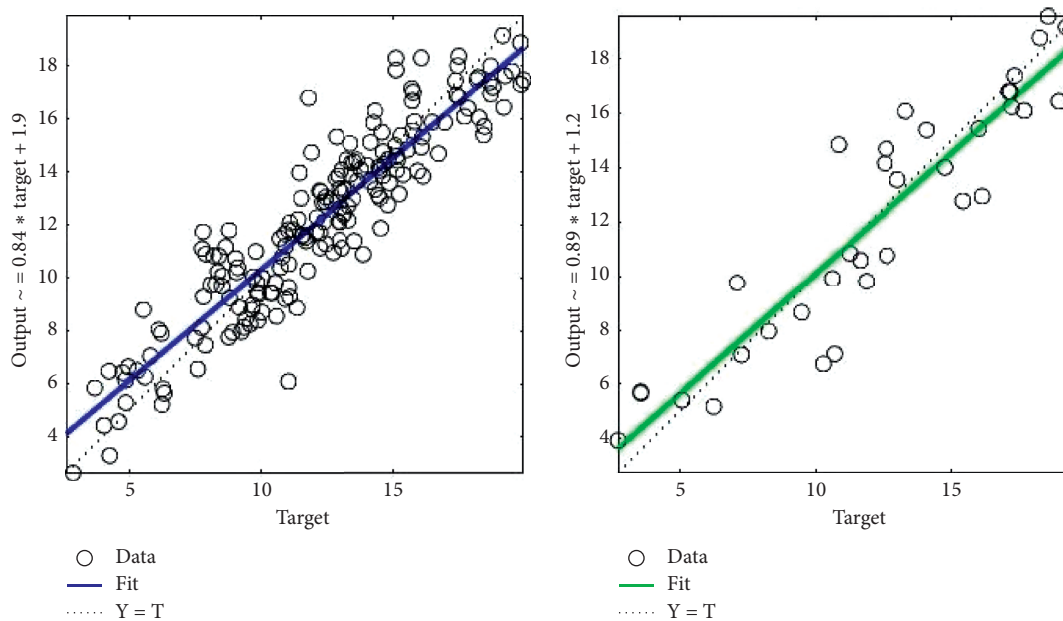

FIGURE 4: Correlations between GSR estimated $(MJm^{-2}d^{-1})$ by ANN and measured GSR $(MJm^{-2}d^{-1})$. The number of hidden neurons = 10.

TABLE 3: Statistical indicators for ANN models.

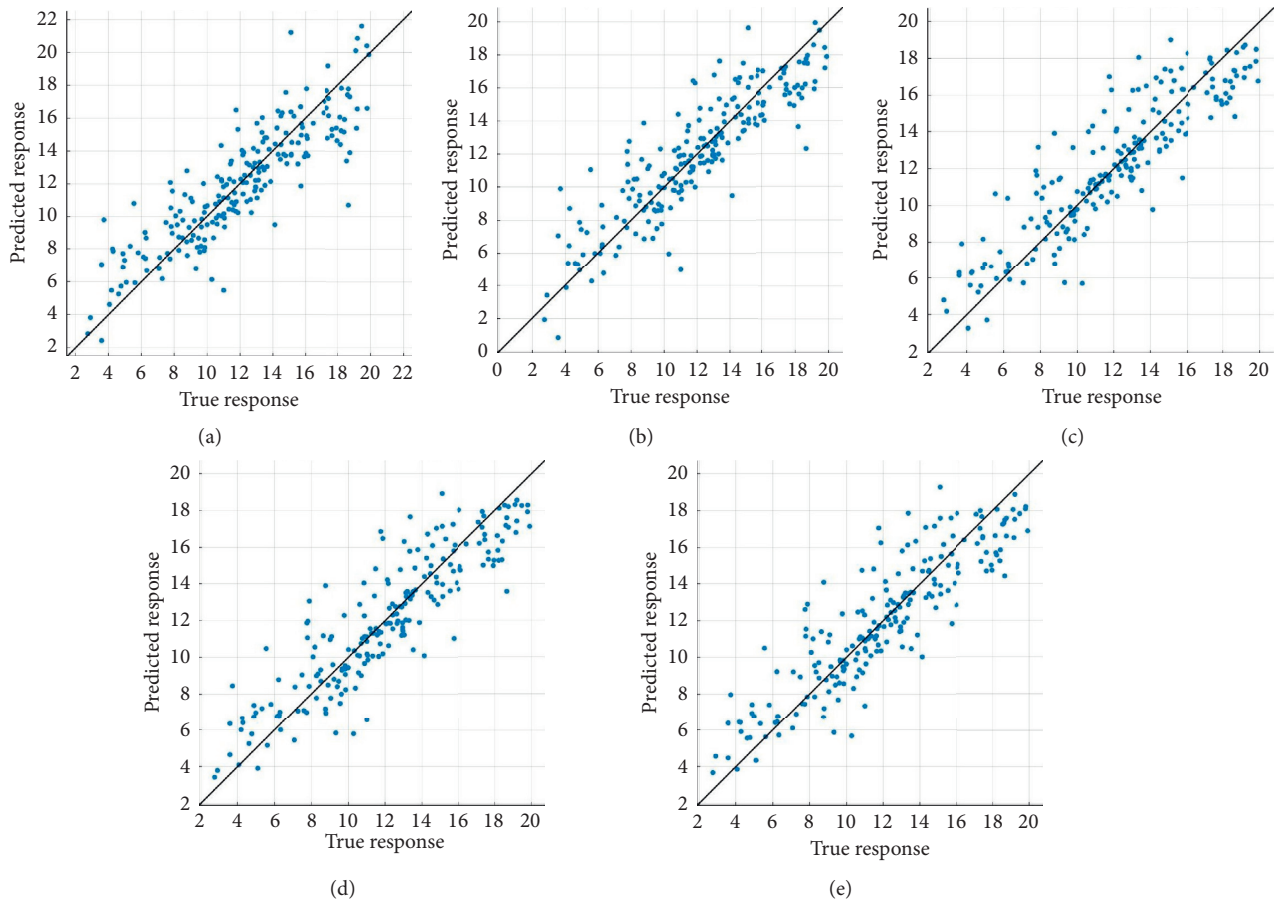| Number of hidden neurons	RMSE (MJm^{-2}d^{-1})		R^2	
	Training	Validation	Training	Validation
3 (ANN1)	2.0001	1.8547	0.7610	0.7240
4 (ANN2)	1.8221	1.8028	0.7903	0.7911
6 (ANN3)	1.6823	2.1954	0.8341	0.7397
8 (ANN4)	2.1564	1.6000	0.8000	0.8479
10 (ANN5)	1.4967	1.8221	0.8485	0.8507

(a)

(b)

(c)

(d)

(e)

FIGURE 5: Scatter plots of measured daily GSR (MJm^{-2}d^{-1}) versus estimated daily GSR (MJm^{-2}d^{-1}) using different ML models with 5-fold cross-validation. Medium Gaussian SVM shows a strong correlation between estimated and measured value. (a) Linear regression. (b) Stepwise linear regression. (c) Medium Gaussian SVM. (d) GPR matern 5/2. (e) GPR exponential.

TABLE 4: Statistical indicators for ML models.

Models	Cross-validation: 5-folds	
	RMSE (MJm^{-2}d^{-1})	R^2
Linear regression	2.010	0.76
Stepwise linear regression	1.940	0.77
Medium Gaussian SVM	1.867	0.79
GPR matern 5/2	1.870	0.79
GPR exponential	1.891	0.79

On the contrary, the cross-validation has little or no effect on the performance of the linear regression and stepwise linear regression model. In the model development, most of these ML models perform extremely well in comparison with temperature-based empirical models. On the contrary, the ANN5 model accounts for greater variance in the training data in comparison with that in ML models.

3.4. Performance Comparison of Empirical and ML Models.
Model assessment is carried out on the 15% unseen data after the model development. RMSE and R^2 of empirical and ML models on the test data are illustrated in Figures 6 and 7, respectively. All of the temperature-based empirical models perform reasonably well on the test data. Amongst the empirical models, the model developed by Fan et al. outperforms other models with an R^2 of 0.7498 and RMSE of 2.0162 MJm^{-2}d^{-1}. A comparison of statistical indicators shows that R^2 and RMSE sometimes follow a different trend.
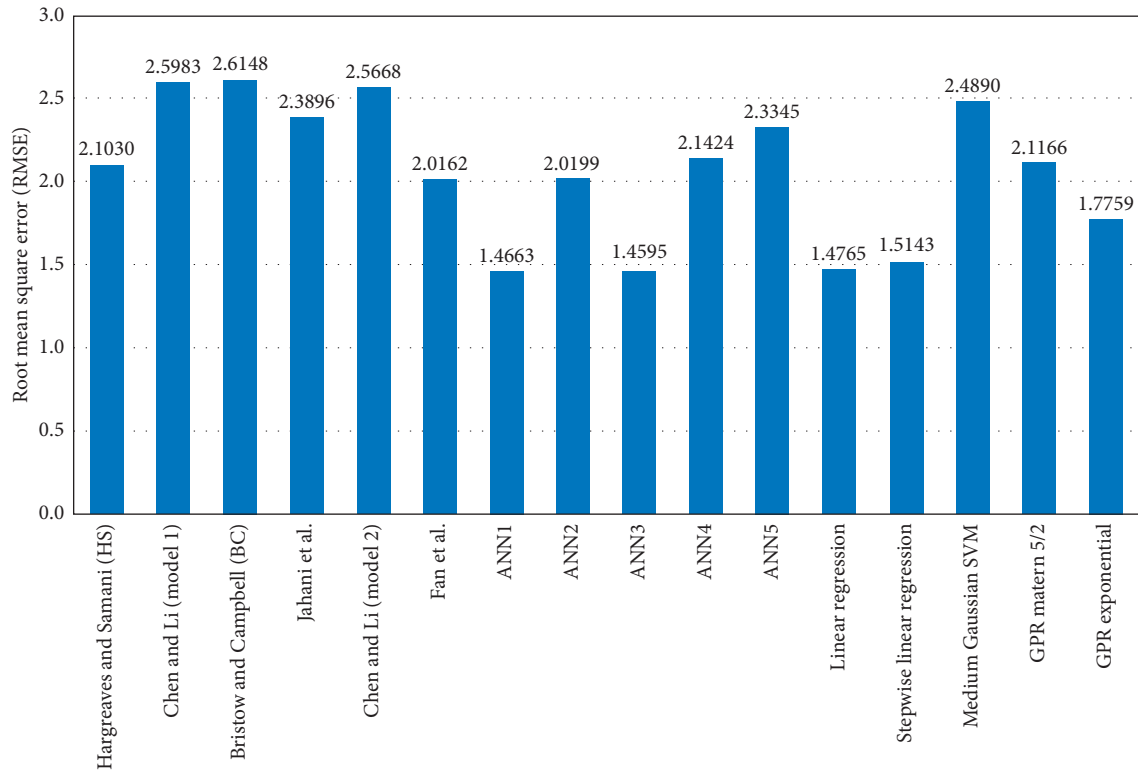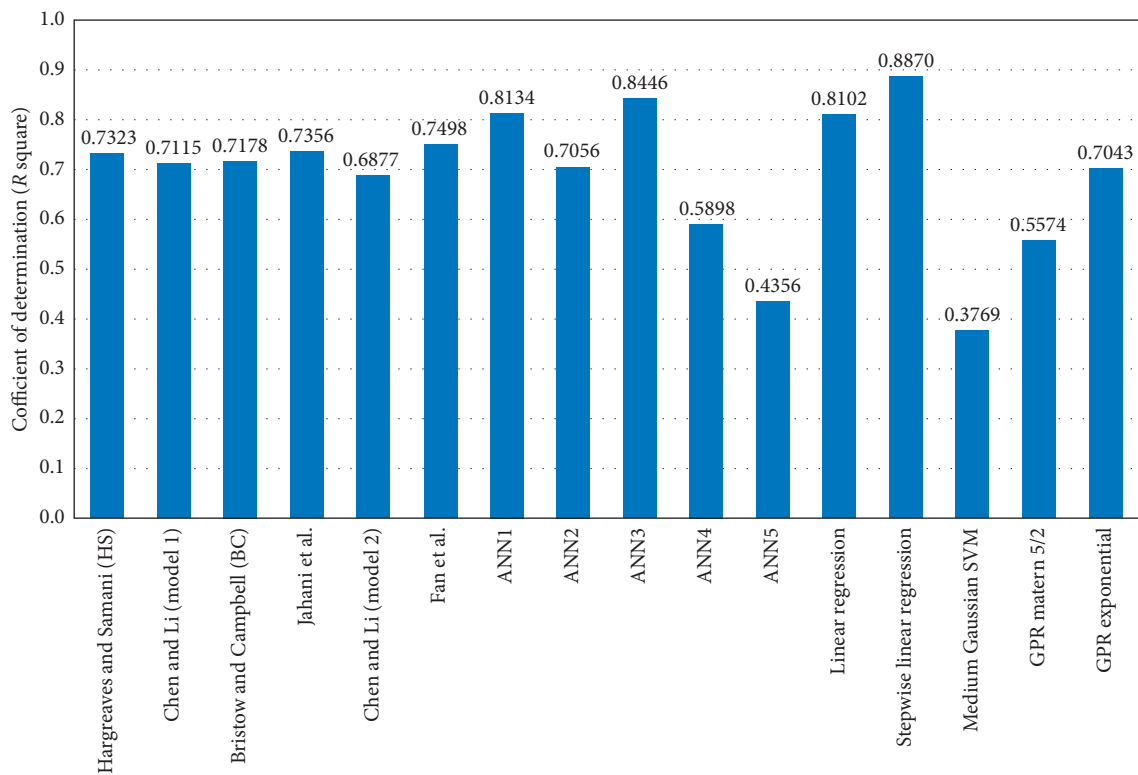
Figure 6: RMSE of empirical and ML models on test data.

Figure 7: R^2 of empirical and ML models on test data.

HS model ranks second among empirical models in terms of RMSE while exhibiting an R^2 of 0.7323. Similarly, the model proposed by Jahani et al. also performs reasonably well with an R^2 of 0.7356 and RMSE of 2.3896 MJm^{-2}d^{-1}.

The performance of ANN on the test data is not consistent with the performance during the model development. Amongst the ANN models, ANN3 performs better than other models with an R^2 of 0.8446 and RMSE of

$1.4595\,\mathrm{MJm^{-2}d^{-1}}$. ANN1 ranks second with an R^2 of 0.8134 and RMSE of $1.4663\,\mathrm{MJm^{-2}d^{-1}}$. The performance of ANN4 and ANN5 degrades substantially in the model assessment in comparison with the model development. The overfitting of the training data during model development is the primary reason for the degradation of the performance.

Similar to the ANN models, the performance of other ML models on the test data is also not consistent with the performance during the model development. A comparison of statistical indicators shows that R^2 and RMSE sometimes follow a different trend. Concerning R^2, stepwise linear regression performs better than other ML models with an R^2 of 0.8870 and RMSE of $1.5143\,\mathrm{MJm^{-2}d^{-1}}$. Likewise, concerning RMSE, linear regression learner performs better than other ML models with R^2 of 0.8102 and RMSE of $1.4765\,\mathrm{MJm^{-2}d^{-1}}$. The performance of medium Gaussian SVM, GPR matern 5/2, and GPR exponential degrades substantially in the model assessment in comparison with the model development. The overfitting of the training data during model development is the primary reason for the degradation of the performance.

4. Conclusion

The present study analyzes and evaluates six different temperature-based empirical models, ANN, and other five different ML models to estimate daily GSR at Biratnagar Airport. Initially, six different temperature-based empirical models with ambient temperature range and daily average temperature as inputs are calibrated. In the model assessment, the model developed by Fan et al. performs better in comparison with other temperature-based models with anR^2 of 0.7498 and RMSE of $2.0162\,\mathrm{MJm^{-2}d^{-1}}$. The performance of ANN with a higher number of neurons in the hidden layer degrades substantially in the model assessment because a large number of neurons tend to overfit the dataset. In the model assessment, ANN3 performs better than other ANN models with an R^2 of 0.8446 and RMSE of $1.4595\,\mathrm{MJm^{-2}d^{-1}}$. Similarly, ANN1 ranks second with an R^2 of 0.8134 and RMSE of $1.4663\,\mathrm{MJm^{-2}d^{-1}}$. Five different ML models available in the MATLAB Regression Learner are analyzed and evaluated to determine the best performing ML model. The performance of medium Gaussian SVM, GPR matern 5/2, and GPR exponential degrades substantially in the model assessment because of the overfitting of the training data during the model development. Concerning R^2, stepwise linear regression performs better than other ML models with an R^2 of 0.8870 and RMSE of $1.5143\,\mathrm{MJm^{-2}d^{-1}}$. Likewise, concerning RMSE, linear regression learner performs better than other ML models with an R^2 of 0.8102 and RMSE of $1.4765\,\mathrm{MJm^{-2}d^{-1}}$.

Considering the generalization capability of temperature-based empirical models, the model proposed by Fan et al. is recommended to estimate daily GSR in the region where only data pertaining to ambient temperature are available. For the regions where data about sunshine duration, maximum and minimum ambient temperature, precipitation, and relative humidity are available, a more robust ANN3 and the stepwise linear regression model are recommended to estimate daily GSR.

Nomenclature

GSR:	Global solar radiation
MLP:	Multilayer perceptron
SVM:	Support vector machine
ANN:	Artificial neural network
DHM:	Department of Hydrology and Meteorology
ANFIS:	Adaptive neurofuzzy inference system
GP:	Genetic programming
RF:	Random forest
RBFN:	Radial basis function network
GPR:	Gaussian process regression
R_s:	Measured global solar radiation
R_a:	Extraterrestrial solar radiation
H_{sc}:	Solar constant ($1367\,\mathrm{W/m^2}$)
n:	Day of the year
N:	Day length
ϕ:	Latitude (rad)
δ:	Solar declination angle (rad)
ω:	Sunshine hour angle
T_{max}:	Maximum ambient temperature
T_{min}:	Minimum ambient temperature
ΔT:	Ambient temperature range
T_a:	Average daily temperature.

Data Availability

The data used to support the findings of this study are supplied by the Department of Hydrology and Meteorology, Government of Nepal, and are included within the supplementary information file.

Conflicts of Interest

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

Acknowledgments

Supplementary Materials

Supplementary 1: raw data of daily global solar radiation, sunshine duration, maximum and minimum ambient temperature, relative humidity, and rainfall of Biratnagar Airport, Morang, Nepal, used for the study are included in the supplementary information file. Supplementary 2: Solar Radiation file contains average solar radiation data measured per second for a particular day. So, daily GSR =

average GSR data measured per second × sunshine duration × 3600. Supplementary 3: Sunshine Duration file contains daily sunshine duration data in hr. Supplementary 4: $T_{\max}$ and $T_{\min}$ file contains daily maximum and minimum ambient temperature data in ^{0}C. Supplementary 5: Rainfall file contains daily rainfall data in millimeters (mm). Supplementary 6: RH file contains data of relative humidity computed a couple of times a day. (*Supplementary Materials*)

References

[1] J. Hansen, P. Kharecha, M. Sato et al., "Assessing "dangerous climate change": required reduction of carbon emissions to protect young people, future generations and nature," *PLoS One*, vol. 8, p. 12, Article ID e81648, 2013.

[2] WECS, "Energy sector synopsis report," 2010.

[3] B. Jahani, Y. Dinpashoh, and A. Raisi Nafchi, "Evaluation and development of empirical models for estimating daily solar radiation," *Renewable and Sustainable Energy Reviews*, vol. 73, pp. 878–891, 2017.

[4] M. A. Behrang, E. Assareh, A. Ghanbarzadeh, and A. R. Noghrehabadi, "The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data," *Solar Energy*, vol. 84, no. 8, pp. 1468–1480, 2010.

[5] B. Amrouche and X. le Pivert, "Artificial neural network based daily local forecasting for global solar radiation," *Applied Energy*, vol. 130, pp. 333–341, Oct. 2014.

[6] J. Fan, B. Chen, L. Wu, F. Zhang, X. Lu, and Y. Xiang, "Evaluation and development of temperature-based empirical models for estimating daily global solar radiation in humid regions," *Energy*, vol. 144, pp. 903–914, 2018.

[7] J. L. Chen, "Empirical models for estimating monthly global solar radiation: a most comprehensive review and comparative case study in China," *Renewable and Sustainable Energy Reviews*, vol. 108, pp. 91–111, 2019.

[8] C. Voyant, C. Darras, M. Muselli, C. Paoli, M.-L. Nivet, and P. Poggi, "Bayesian rules and stochastic models for high accuracy prediction of solar radiation," *Applied Energy*, vol. 114, pp. 218–226, 2014.

[9] N. Lu, J. Qin, K. Yang, and J. Sun, "A simple and efficient algorithm to estimate daily global solar radiation from geostationary satellite data," *Energy*, vol. 36, no. 5, pp. 3179–3188, 2011.

[10] F. Besharat, A. A. Dehghan, and A. R. Faghih, "Empirical models for estimating global solar radiation: a review and case study," *Renewable and Sustainable Energy Reviews*, vol. 21, pp. 798–821, 2013.

[11] S. R. Bajracharya, M. S. Shrestha, and A. B. Shrestha, "Assessment of high-resolution satellite rainfall estimation products in a streamflow model for flood prediction in the Bagmati basin, Nepal," *Journal of Flood Risk Management*, vol. 10, no. 1, pp. 5–16, 2017.

[12] K. Bayrakçı, H. Cenk, C. Demircan, and K. Ali, "The development of empirical models for estimating global solar radiation on horizontal surface: a case study," *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2771–2782, 2018.

[13] G. E. Hassan, M. E. Youssef, Z. E. Mohamed, M. A. Ali, and A. A. Hanafy, "New temperature-based models for predicting global solar radiation," *Applied Energy*, vol. 179, pp. 437–450, 2016.

[14] R. Kostić and J. Mikulović, "The empirical models for estimating solar insolation in Serbia by using meteorological data on cloudiness," *Renewable Energy*, vol. 114, pp. 1281–1293, 2017.

[15] J. A. Martínez-Lozano, F. Tena, J. E. Onrubia, and J. de la Rubia, "The historical evolution of the Ångström formula and its modifications: review and bibliography," *Agricultural and Forest Meteorology*, vol. 33, no. 2-3, pp. 109–128, 1984.

[16] J. Awasthi and K. N. Poudyal, "Estimation of global solar radiation using empirical model on meteorological parameters at Simara Airport, Bara, Nepal," *Journal of the Institute of Engineering*, vol. 14, no. 1, pp. 143–150, 2018.

[17] R. C. Srivastava and H. Pandey, "Estimating angstrom-prescott coefficients for india and developing a correlation between sunshine hours and global solar radiation for india," *ISRN Renewable Energy*, vol. 14, pp. 1–7, 2013.

[18] G. P. Podestá, L. Núñez, C. A. Villanueva, and M. A. Skansi, "Estimating daily solar radiation in the Argentine Pampas," *Agricultural and Forest Meteorology*, vol. 123, no. 1-2, pp. 41–53, 2004.

[19] M. Trnka, Z. Žalud, J. Eitzinger, and M. Dubrovský, "Global solar radiation in Central European lowlands estimated by various empirical formulae," *Agricultural and Forest Meteorology*, vol. 131, no. 1-2, pp. 54–76, 2005.

[20] M. S. Mecibah, T. E. Boukelia, R. Tahtah, and K. Gairaa, "Introducing the best model for estimation the monthly mean daily global solar radiation on a horizontal surface (Case study: Algeria)," *Renewable and Sustainable Energy Reviews*, vol. 36, pp. 194–202, 2014.

[21] G. H. Hargreaves and Z. A. Samani, "Estimating potential evapotranspiration," *Journal of Irrigation and Drainage Division*, vol. 108, no. 3, pp. 225–230, 1982.

[22] D. G. Goodin, J. M. S. Hutchinson, R. L. Vanderlip, M. C. Knapp, and D. G. Goodin, "Estimating solar irradiance for crop modeling using daily air temperature data," *Agronomy Journal*, vol. 91, no. 5, pp. 845–851, 1999.

[23] A. P. Oliveira, J. Soares, M. Z. Božnar, P. Mlakar, and J. F. Escobedo, "An application of Neural Network Technique to correct the dome temperature effects on Pyrgeometer measurements," *Journal of Atmospheric and Oceanic Technology*, vol. 23, no. 1, pp. 80–89, 2006.

[24] C. Voyant, "Machine learning methods for solar radiation forecasting: a review," *Renewable Energy*, vol. 105, pp. 569–582, 2017.

[25] S. Belaid and A. Mellit, "Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate," *Energy Conversion and Management*, vol. 118, pp. 105–118, 2016.

[26] A. Mellit, "Department of environment," 2020.

[27] A. Mellit, "Climate-data," 2020.

[28] J.-L. Chen and G.-S. Li, "Estimation of monthly average daily solar radiation from measured meteorological data in Yangtze River Basin in China," *International Journal of Climatology*, vol. 33, no. 2, pp. 487–498, 2013.

[29] K. L. Bristow and G. S. Campbell, "On the relationship between incoming solar radiation and daily maximum and minimum temperature," *Agricultural and Forest Meteorology*, vol. 31, no. 2, pp. 159–166, 1984.

[30] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes*, Wiley, London, UK, 2013.

[31] P. I. Cooper, "The absorption of radiation in solar stills," *Solar Energy*, vol. 12, no. 3, pp. 333–346, 1969.

[32] J. C. Lam, K. K. W. Wan, and L. Yang, "Solar radiation modelling using ANNs for different climates in China," *Energy Conversion and Management*, vol. 49, no. 5, pp. 1080–1090, 2008.

[33] S. Rehman and M. Mohandes, "Artificial neural network estimation of global solar radiation using air temperature and relative humidity," *Energy Policy*, vol. 36, no. 2, pp. 571–576, 2008.

[34] S. K. Shevade, S. S. Keerthi, C. Bhattacharyya, and K. R. K. Murthy, "Improvements to the SMO algorithm for SVM regression," 2000.

[35] The Mathworks Inc., *Neural Net Fitting*, The Mathworks Inc., Natick, MA, USA, 2019.

[36] The Mathworks Inc., *Regression Learner App*, The Mathworks Inc., Natick, MA, USA, 2019.