Global Solar Radiation Forecasting using Artificial Neural Network and Support Vector Machine

Emmanuel Philibus¹, Roselina Sallehuddin², Yusliza Yussof² and Lizawati Mi Yusuf²

¹Department of Computer Science, Kaduna State College of Education, Gidan Waya, (Kaduna), Nigeria.
²Department of Computer Science, School of Computing, Universiti Teknologi Malaysia, Skudai, (Johor), Malaysia.

Email: emmanuel-1984@graduate.utm.my

Abstract. Global solar radiation (GSoR) forecasting involves predicting future energy from the sun based on past and present data. Literature reveals that not all meteorological stations record solar radiation, some equipments are faulty, and are not available in every location due to high cost. Hence, the need to predict and forecast using predictors such as land surface temperature (LST). Satellite data when were used to complement ground-based stations have been yielding good results. Different artificial intelligence (AI) methods such as Support Vector Machine (SVM) and Artificial Neural Network (ANN) present different forecasting performances. Motivated by existing literature-related contradictions on the performance superiority of ANN and SVM in GSoR forecasting, the two techniques were compared based on several statistical tests. Experimental results show that ANN outperformed SVM by 2.9864% accuracy, making it superior in the forecast of GSoR.

1. Introduction

Global solar radiation (GSoR) forecasting is the process of predicting future energy from the sun based on past and present data. Literatures reveal that not all meteorological stations record solar radiation, some equipment are faulty thereby given inaccurate measurements. Other challenges include non-availability of measuring equipment in every location due to high cost. Hence, the need to predict and forecast using satellite-derived predictors such as land surface temperature (LST). Like other time series, GSoR forecasting helps in mitigating risks in decision making usually made by government policymakers, energy experts, etc., when planning to regulate power usage, optimise energy balance and environmental effects, and/or increase photovoltaic (PV) plants [1].

Artificial intelligence (AI) techniques are preferred over statistical methods in forecasting GSoR due to their ability to extract complex nonlinear relationships in data. Common AI techniques in use include support vector machine (SVM), extreme gradient boosting (XGBoost), artificial neural network (ANN), random forest and gradient boosted regression trees (GBRT) [1-10] among others. Despite the use of these AI techniques in forecasting GSoR, however, a contradiction exists in reported literature on the superiority of the popular techniques - ANN and SVM [2,11-12]. The result reported by Quej et al. [13] disagrees with that of Hamilton et al. [2]. Motivated by this rationale, this study compares ANN and SVM in forecasting GSoR.
2. Materials and Method

2.1 Study Sites

The study site comprises three locations from Queensland, Australia (Table 1). Australia was selected because of its potential for solar, being characterised by high insolation, low cloud coverage, and low rainfall. The three study sites are: St. George, St. George Post Office, and St. George Airport. The model is trained using the first two sites and forecasted using the third location (Table 1).

Table 1. Characteristics of study sites with monthly average satellite-derived land-surface temperature (LST) and solar radiation (GSoR) for training (2008 – 2018) and forecasting (2019).

| Training and Testing Sites | Location (Lon, Lat) | Altitude (m) | LST (K) | GSoR (MJ m$^{-2}$ month$^{-1}$) |
|---------------------------|---------------------|--------------|---------|---------------------------------|
|                           |                     |              | Training 2008-2018 | Forecasting 2019 | Training 2008-2018 | Forecasting 2019 |
| St. George; ID 43053      | 28.0533°S, 148.5633°E | 200.0 m     | 304.07   | 19.50                           |
| St. George Post Office ID | 28.0361°S, 148.5814°E | 200.8 m     | 301.28   | 19.10                           |
| Cross-Validation: St.     | 28.0489°S, 148.5942°E | 198.5 m     | 305.38   | 19.00                           |
| George Airport; ID 43109  |                     |              |          |                                 |

Land surface temperature (LST) data were obtained from National Aeronautics and Space Administration (NASA) built satellite, and global solar radiation (GSoR) from Scientific Information for Land Owners (SILO) for twelve years – eleven years for model development and one year for cross-validation. The primary predictor is the LST while the target variable is GSoR. Other accompanying site-related features were the altitude, latitude, longitude, and months (which is the solar periodicity).

2.2 Data pre-processing and normalisation

In this study, the data were checked for availability of records containing missing and zero values and were replaced with the mean of the non-zero values as a rationale value because deleting such records is not recommended. Then, min-max normalization was applied to scale the data in the range 0 to 1. Scaling ensures that variables (attributes) with large values do not dominate smaller ones to bias the result.

2.3 Evaluation metrics

Although the best performance metrics in forecasting are MAPE and accuracy, this study also included other indicators such as MSE, $R^2$, and RMSE. In this study, the $R^2$, RMSE, and MAPE are used at the training stage to determine the best configuration’s performance. At the final forecast stages, the RMSE, MAPE, and accuracy are used to compare and determine the best-implemented model.

2.4 Artificial Neural Network

The ANN was configured with LST obtained from the moderate resolution imaging spectroradiometer (MODIS) of the NASA satellite, which is used as the main predictor together with study-site parameters such as longitude, latitude, altitude, and solar period (months) as the model’s regression inputs to forecast the target variable, GSoR (Figure 1). To obtain the value of neurons at the hidden layer that gives a competent network, a rule of thumb was applied. It suggests that, to prevent over-fitting, the neurons in the hidden layer ($N_h$) be kept below the value:

$$N_h = \frac{N_s}{(\alpha \times (N_t + N_0))}$$  (1)
where:

\( N_i \) is the value of input neurons; \( N_o \) is the value of output neurons; \( N_s \) is the value of samples in the training data set; \( \alpha \) is an arbitrary scaling factor usually 2-10; \( N_i = 5, N_o = 1, N_s = 264 \), and \( \alpha = 2 \) (being most common value). Hence, \( N_h \) is 22 (using Equation 1).

Figure 1. ANN Model Architecture.

Figure 2. SVM Model Architecture.

2.5 Support Vector Machine
The same regressors were used to configure a regression version of support vector machine (SVM) which uses a linear model to create a hyperplane to implement nonlinear class boundaries in a high-dimensional feature space through some nonlinear mapping input vectors (Figure 2). Regression problems in SVM can be applied to include distance measurement by introducing the alternative loss function [18]. Support vector regression (SVR) model which is the regression form of SVM was setup with a cross validation value of 30. The RBF kernel was used being common in the literature. The output provides the total MSE and \( R^2 \). To get the best of the experiment, the kernel was first configured with default parameters \( C, \gamma, \varepsilon \), to be 1, 0.2, and 0.1, respectively. Subsequently, a grid search tuning technique was applied to obtain optimal value for cost and gamma (\( C, \gamma \)).

3. Results and Discussions

3.1 Artificial Neural Network Implementation
The maximum number of neurons for the hidden layer (\( N_h \)) were found to be 22. Therefore, trials were made for the values of the hidden neurons from 2 to 22. Table 2 shows the two models of ANN using Tanh and logistic activation functions.

| Model Trials | Act. Func. | Model Arch. (Input-Hidden-Output) | Training Coef. Of Det. \( (r^2) \) | Forecast Coef. Of Det. \( (r^2) \) | MSE | RMSE | MAPE |
|--------------|------------|----------------------------------|----------------------------------|----------------------------------|-----|------|------|
| M2           | Tanh       | 5 – 3 – 1                         | 0.8593                           | 0.9040                           | 0.0073 | 0.0853 | 0.1579 |
| M4           | Tanh       | 5 – 5 – 1                         | 0.8719                           | 0.8853                           | 0.0089 | 0.0945 | 0.2222 |
| M22          | logistic   | 5 – 2 – 1                         | 0.8564                           | 0.9024                           | 0.0078 | 0.0883 | 0.1431 |
| M25          | logistic   | 5 – 5 – 1                         | 0.8719                           | 0.9079                           | 0.0313 | 0.1769 | 0.2180 |

The best model (Table 2) is determined by the lowest RMSE, MAPE, and highest \( R^2 \) [1]. Therefore, the best ANN model is M22. The result of the training is shown in Table 2 (highlighted in bold) and figure 3. The model recorded an \( R^2 \) of 0.8564 during training and 0.9024 at forecast. The accuracy of the training and forecast were computed to be 91.7507% and 93.9188%, respectively.
In modelling, forecast performance is the most important, and priority is giving to the lowest value of RMSE or MAPE. The closer the value of RMSE to zero, the better the model and vice-versa to 1. The RMSE in M22 (i.e., 0.0883) was accepted being the lowest and justified by the value of MAPE (0.1431). Consequently, Figure 5 presents a one-year (2019) forecast using M22 (Table 2).

3.2 Support Vector Machine Implementation
The grid search tuning yielded the best performance parameters for cost and gamma as 1.04 and 0.15, respectively. Table 3 shows the SVM forecasting performance using grid search tuning parameter.

The tuned parameter values were used to train the SVM (figure 4). Furthermore, GSoR was forecasted using this tuned RBF-SVM model for one-year (2019). The result of the prediction yielded an R^2 of 0.9112 at an RMSE of 0.1014 (Table 3, Figure 5).

3.3 Comparison between ANN and SVM
The summary of the two techniques is presented in Table 4 - putting the best models’ values together using the R^2, RMSE, MAPE, and accuracy (in forecast).

To compare the forecasted and actual monthly GSoR at the cross-validation period (2019), Figures 5 and 6 plots the models’ forecasts and errors, respectively. Note that, the errors were calculated by taking the absolute values of the difference between actual and forecasted GSoR (i.e., |GSoRA - GSoRF|), each month. Notably, the SVM model recorded the highest error magnitude (MAPE=9.0676MJm^-2month^-1) which led to its low accuracy in forecast. High MAPE produces low accuracy and vice versa. The forecast errors are presented in Figure 6.
The result in Table 4, Figure 5, and Figure 6, presented ANN as a superior model over SVM having recorded a higher forecast accuracy of 93.9188% at a lower MAPE of 6.0812MJm⁻²month⁻¹.

4. Conclusions

Two separate analyses were carried out and discussed. Data extraction processes as it relates to MODIS and SILO was the first task. Secondly, SVM and ANN models were implemented to forecast GSoR in comparison. The global nature of the models is in fitting one model for the training locations instead of one for each. The result of ANN’s forecast was an $R^2$ of 0.9024, a MAPE of 6.0812MJm⁻²month⁻¹, and an accuracy of 93.9188%. On the other hand, the resultant forecast from the grid search tuning of SVM gave an $R^2$ of 0.9112, a MAPE of 9.0676MJm⁻²month⁻¹, and an accuracy of 90.9324%. Conclusively, ANN outperformed SVM by 2.9864% accuracy, making it superior in the forecast of GSoR.

References

[1] Deo, R C and Şahin, M 2017 Ren. and Sust. Ener. Rev. 72, 828-848.
[2] Yadav, A K, & Chandel, S 2014 Ren. and Sust. Ener. Rev. 33, 772-781.
[3] Hamilton, C, Potter, W, Hoogenboom, G, McClendon, R, & Hobbs, W 2015 Int. J. Comput. Control Quant. Inf. Eng. 9(5), 656-661.
[4] McGovern, A., Gagne, D. J., Basara, J., Hamill, T. M., & Margolin, D. 2015 Bulletin of the Amer. Meteo. Soc. 96(8), 1388-1395.
[5] Fan, J, Wang, X, Wu, L, Zhang, F, Bai, H, Lu, X, & Xiang, Y 2018 Ener. Conv. and manag., 156, 618-625.
[6] Sobri, S, Koohi-Kamali, S, & Rahim, N A 2018 Ener. Conv. and manag., 156, 459-497.
[7] Tealab, A 2018 Fut. Comp. and Infor. J.
[8] Zendehboudi, A Baseer, M, & Saidur, R 2018 J. of Clean. Prod., 199, 272-285.
[9] Benali, L, Notton, G, Fouilloy, A, Voyant, C, & Dizene, R 2019 Ren. Ener. 132, 871-884.
[10] Elsheikh, A H, Sharshir, S W, Elaziz, M A, Kabeel, A, Guilan, W, & Haiou, Z 2019 Sol. Ener. 180, 622-639.
[11] Ren, Y, Suganthan, P, & Srikanth, N 2015 Ren. and Sust. Ener. Rev. 50, 82-91.
[12] Ramli, M A, Twaha, S, & Al-Turki, Y A 2015 Ener. Conv. and manag. 105, 442-452.
[13] Quej, V H, Almorox, J, Amenlao, J A, & Saito, L 2017 J. of Atm. and Sol.-Terr. Phy. 155, 62-70.

Acknowledgement

The authors wish to thank the Ministry of Higher Education, Malaysia, for funding this study under the Fundamental Research Grant Scheme FRGS/1/2019/ICT02/UTM/02/13. The authors would like to thank Research Management Centre (RMC) Universiti Teknologi Malaysia, and ALI@S for their support in research activities.