7th European Conference on Industrial Engineering and Operations Management
Augsburg, Germany, July 16-18, 2024
Publisher: IEOM Society International, USA DOI: 10.46254/EU07.20240161
Published: July 16, 2024

Estimation of Global Horizontal Irradiance Using several Deep-learning Approaches

Fathy Alkhatib, Maher Maalouf and Mohammad Omar
Department of Industrial and Systems Engineering, Khalifa University
Abu Dhabi, United Arab Emirates
100062700@ku.ac.ae, maalouf@ku.ac.ae, mohammed.omar@ku.ac.ae

Abstract

The design of solar energy systems poses significant challenges, encompassing issues related to infrastructure, installation costs, and land availability and selection. In the practical implementation of such projects, location selection is critically dependent on numerous vital criteria, one of which is the estimation of received solar radiation. For instance, solar radiation estimation is pivotal for designing proper energy storage systems, balancing load, and managing photovoltaic and thermal systems. Machine and deep learning approaches have been commonly utilized in solar radiation forecasting and estimation. This paper deploys multiple deep learning techniques, including feed-forward neural network, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN) to estimate the amount of Global Horizontal Irradiance (GHI) received in the city of Tustin, California, using multiple metrological factors. California was selected for this study because it is characterized by receiving high amounts of solar radiation and large amounts of solar energy installations and projects. The paper compares the three examined deep learning methods using the commonly used evaluation metric, the Root Mean Squared Error (RMSE). The findings indicate that the LSTM approach is the most effective deep learning technique for estimating the hourly GHI.

Keywords
Deep learning, Forecasting, Global horizontal irradiance, Solar radiation.

1. Introduction

In the modern era, energy plays a huge role in the industrialization and development of societies. Burning fossil fuels is the main energy source and significantly meets global energy loads (Daut et al., 2011). Fossil fuel provisions are finite, and their use has detrimental effects on the environment (Daut et al., 2011). They release many toxicants that disrupt the planet's ecological life and lead to various human ailments. Hence, it is crucial to reduce the dependence on fossil-based fuels and replace them with renewable and sustainable energy sources (Ibrahim et al., 2012). Solar energy has seen a sharp increase in popularity in recent years because of its abundance, declining cost compared to other sustainable energy sources, and ease of conversion into other useful forms of electrical energy. Designing solar energy conversion systems and several agricultural applications requires good solar radiation estimation. However, most weather stations need more modern and accurate solar radiation measurement devices, yet they have accurate temperature and wind speed sensors (Ouammi et al., 2010). Additionally, many developing countries that rely on agriculture lack access to reliable sources for solar radiation measurements (Shahsavar and Akbari, 2018). Thus, researchers tried to develop accurate solar radiation estimation models. The estimation models can be categorized into three classes: empirical models, statistical models, and machine learning models. Empirical methods include linear and nonlinear regression models. These models are easy to use and implement but don't provide the highest accuracy (Gomila, 2021). Statistical models encompass time-series and autoregressive integrated moving average (ARIMA) models (Kaur and Rakshit, 2020). Lastly, machine and deep learning models and algorithms, such as support vector machines and neural networks, provide the best possible performance compared to other models (Kumari and Toshniwal, 2021; Gürel et al., 2023). In addition, machine learning models perform reasonably well under various operating conditions (Sivamadhavi, V. and Selvaraj, 2012). In this paper, deep learning models such as Artificial

Neural Networks (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) are used to estimate the global horizontal irradiance (GHI) based on historical GHI measurements. GHI is the total amount of solar radiation received on a specific horizontal surface per area unit (Gbémou et al., 2021). The study is conducted in the city of Tustin, California, because the city is considered a hub for solar energy installation projects, owing to its suitable climate and weather conditions.

This paper will be structured as follows: the implemented method and data description are presented in Section 2, whereas the results and discussion are expressed in Section 3. Lastly, Section 4 presents the conclusion of the paper discussing some key findings and implications.

2. Materials and Methods

This paper deploys various deep-learning models to forecast GHI. ANN, LSTM, and CNN are deployed to predict the hourly *GHI* values in Tustin, California. These methods are data intensive; therefore, hourly measurements over one year, 2022 (8760 data points), are used to train and test the models. A rolling window is used in the deployed deep learning techniques with a size of 48, which means the models use the previous 48 hours before making the predictions. Additionally, the number of epochs is ten. This indicates that the rolling window traverses the entire data set ten times throughout the training phase. The number of epochs has been selected because ten is a commonly used number that usually does not result in underfitting or overfitting in the model. The learning rate is also fixed at a commonly used value of 0.01 throughout the entire study.

Table 1. The architectures used for the deployed deep learning methods.

Method	**Architecture**
ANN	Input layer: 128 neurons, each of size (48, 1) Hidden layer 1: 64 neurons Hidden layer 2: 32 neurons Output layer: 1 neuron
LSTM	Input layer: 1 neuron of size (48, 1) LSTM layer: 100 neurons Hidden layer 1: 32 neurons Output layer: 1 neuron
CNN	Input layer: 1 neuron of size (48, 1) Convolutional layer: 64 neurons, with a kernel size of 2 Flatten layer Hidden layer 1: 8 neurons Output layer: 1 neuron

To assess the performance of the developed models, the root mean squared error (RMSE) will be deployed. The previously mentioned performance metric is commonly used in literature for solar irradiance prediction. For instance, the RMSE is widely used because it gives higher weight to large errors, which is crucial in solar irradiance prediction where significant deviations can have a substantial impact on energy production. (Yang et al., 2018).

In the input layer of the ANN and the hidden layers of the LSTM and CNN, the rectified linear unit (ReLU) is selected as an activation function because of its high computational efficiency. The kernel's role in the convolutional layer is to help the network extract more information from the data. This work uses a kernel size of 2 in the CNN, as the data is completely numeric. Table 1 illustrates the architecture of the used ANN, LSTM, and CNN.

Table 2. Used forecasting models

Sr. No.	Method
1	Artificial neural network (ANN)
2	Long short-term memory (LSTM)
3	Convolutional neural network (CNN)

It is vital to note that every method employed in this study is for univariate GHI forecasting, indicating that past GHI values are used to predict future ones and that no additional independent parameters are used to help forecast future GHI. Table 2 summarizes all methods and techniques presented in this project.

The data used in this paper is obtained from the National Solar Radiation Database (NSRDB). The data consists of several meteorological measurements recorded hourly in the year 2021, such as DNI, DHI, air temperature, wind speed, dew point, fill flag, pressure, surface albedo value, and GHI for Tustin, California. This paper reports on the univariate prediction of GHI using the models shown in Table 2. Table 3 summarizes a few descriptive statistics for some features in the data.

Table 3. Descriptive statistics about the data

	GHI (W/m^2)	**Air temperature (C°)**	**Wind speed (m/s)**	**DHI (W/m^2)**
Mean	230	19	2.17	54.76
Std.	315	5.88	1.29	77
Min	0	5.3	0.3	0
25%	0	14.4	1.2	0
50%	0	18.7	1.8	0
75%	450	23	3	94
Max	1064	34.9	9	483

3. Results and Discussion

Three deep-learning methods are investigated in this section. A rolling window function is used in all models, with a window size of 48. The rolling window passes through the entire data set ten times, resulting in an epoch size of ten. The data set is divided into 60% training, 20% cross-validation, and 20% testing. This translates into 5256 data points for training and 1752 data points for each cross-validation and testing data set. The performance is measured based on the cross-validation RMSE.

Initially, a normal dense-layered ANN is implemented with the architecture shown in Table 1. The network produced the least performance compared to all the methods shown in this paper, with a performance measure RMSE = 317.53 after the tenth epoch. This indicates that the ANN was unable to accurately predict the hourly GHI compared to the other implemented deep-learning models. A special kind of recurrent neural network, the LSTM, is implemented to

forecast hourly GHI. Unlike traditional ANNs, LSTM has feedback loops that allow for better learning. However, in comparison to other deep learning methods, like ANN and CNN, LSTM can be computationally expensive. The obtained performance indicator from the LSTM is RMSE = 76.18. For better visualization, the figures below illustrate the actual and predicted GHI values for only the last 100 data points in the testing set. Figure 1 compares the actual GHI and the test-predicted GHI values from the LSTM model.

Furthermore, a CNN model is implemented with a single convolutional layer and a kernel size of 2. The kernel acts as a convolutional filter. A kernel size of 2 enables the network to group each two data points into one value and use these filtered values in the rolling window. This model produced highly improved performance in comparison with the ANN and a comparable performance to the LSTM model. For instance, the resultant RMSE a from this CNN model is 71, resulting in the best-performing model. Figure 2 shows the actual GHI, and the test predicted *GHI* values from the CNN model.

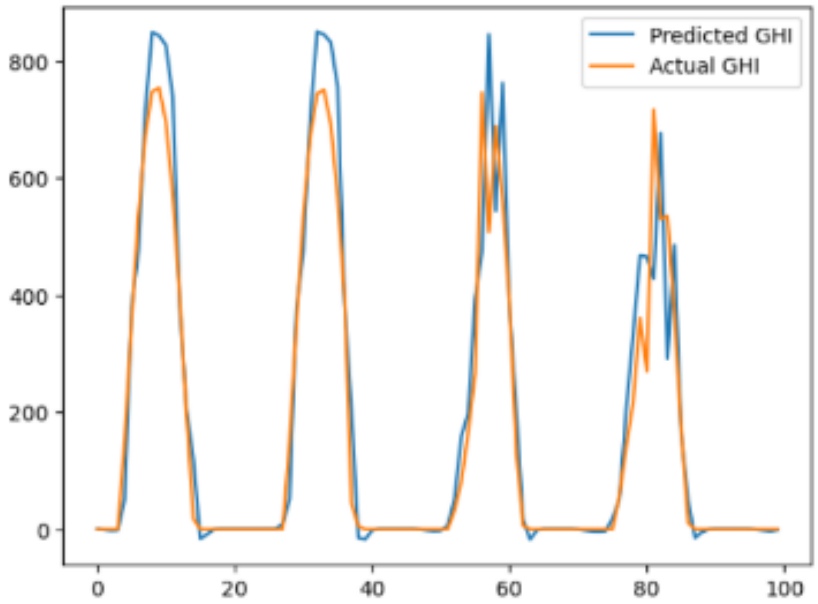

Figure 1. Actual and predicted *GHI* values using LSTM.

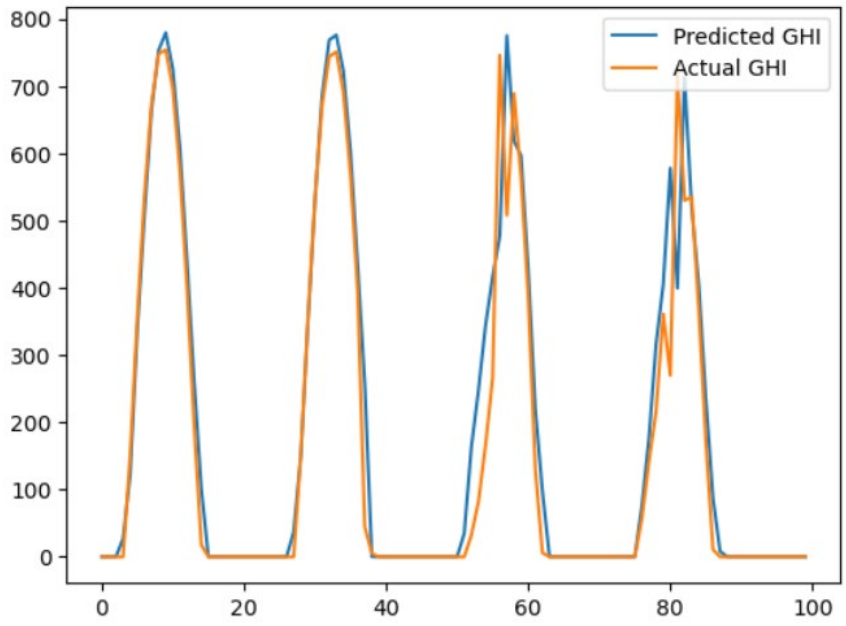

Figure 2. Actual and predicted *GHI* values using CNN.

The findings indicate that LSTM and CNN produced almost identical forecasting performance, while the ANN delivered the weakest performance. This can be mainly attributed to the inherent characteristics of CNN and LSTM that enable them to remember patterns over extended periods of time (Lu et al., 2020). Moreover, CNN can make accurate predictions and estimations because of the filters that allow the network to compute the dilations between various cells (Wibawa et al., 2022). Table 4 summarizes the RMSE for all the deep learning models presented in this paper.

Table 4. Summary of hourly *GHI* forecasting methods

Method	**RMSE**
ANN	317.53
LSTM	76.18
CNN	71.00

4. Conclusion

Energy is a critical sector in any country's economy. In the past few years, governments started opting for more sustainable and renewable energy sources to fight global warming and contribute to achieving the Paris Agreement goals. In parallel, solar energy has become a more accessible and viable option for various countries. However, some developing countries lack modern equipment and methods to help forecast solar radiation, which is pivotal for solar energy installations and projects. Consequently, this paper aims to develop and compare various deep learning univariate models, namely ANN, CNN, and LSTM, to predict the hourly GHI in the city of Tustin, California. The study employed RMSE as a performance measure to facilitate the models' comparison. The findings suggested that ANN produces the lowest performance, as indicated by the high RMSE value of 317.53. In contrast, both CNN and LSTM produced comparable performance with RMSE values of 71 and 76.18, respectively. The findings of the paper can hold significant implications for sustainable and renewable energy advocates interested in solar energy installation projects in the city of Tustin or any other city with comparable climate conditions. Moreover, the paper contributes to the academic literature on solar radiation predictions and estimations by deploying three univariate deep-learning approaches to predict the hourly GHI. Future studies can encompass the exploration of different learning rates and the

impact they have on the performance of the models. Additionally, different deep learning networks architectures can be studied and analyzed. Advanced techniques like adaptive learning rate and the integration of several deep leaning networks are also worth considering.

References

Daut, I., Irwanto, M., Irwan, Y.M., Gomesh, N. and Ahmad, N.S., Combination of Hargreaves method and linear regression as a new method to estimate solar radiation in Perlis, Northern Malaysia, *Solar Energy*, vol. 85, no. 11, pp.2871-2880, 2011.

Gbémou, S., Eynard, J., Thil, S. and Grieu, S., Comparison between Time-and Observation-Based Gaussian Process Regression Models for Global Horizontal Irradiance Forecasting. *Solar MDPI*, vol. 2, no. 4, pp. 445-468, 2022.

Gomila, R., Logistic or linear? Estimating causal effects of experimental treatments on binary outcomes using regression analysis, *Journal of Experimental Psychology: General*, vol. 150, no. 4, p.700, 2021.

Gürel, AE., Ağbulut, Ü., Bakır, H., Ergün, A., and Yıldız, G., A state of art review on estimation of solar radiation with various models. *Heliyon*. Vol. 9, no. 2, 2023.

Ibrahim, S., Daut, I., Irwan, Y.M., Irwanto, M., Gomesh, N. and Farhana, Z., Linear regression model in estimating solar radiation in Perlis, *Energy Procedia*, vol. 18, pp.1402-1412, 2012.

Jia, D., Hua, J., Wang, L., Guo, Y., Guo, H., Wu, P., Liu, M. and Yang, L., Estimations of global horizontal irradiance and direct normal irradiance by using Fengyun-4A satellite data in northern China, *Remote Sensing*, vol.13, no. 4, p.790, 2021.

Kaur, S., and Rakshit, M., Seasonal and periodic Autoregressive time series models used for forecasting analysis of Rainfall data, *International Journal of Advanced Research in Engineering and Technology*, vol. 10, no. 1, p.2019, 2020.

Kumari, P. and Toshniwal, D., Deep learning models for solar irradiance forecasting: A comprehensive review, *Journal of Cleaner Production*, vol. 318, p.128566, 2021.

Lu, W., Li, J., Li, Y., Sun, A. and Wang, J., A CNN-LSTM-based model to forecast stock prices, *Complexity*, pp.1-10, 2020.

Ouammi, A., Dagdougui, H., Sacile, R. and Mimet, A., Monthly and seasonal assessment of wind energy characteristics at four monitored locations in Liguria region (Italy), *Renewable and Sustainable Energy Reviews*, vol. 14, no. 7, pp.1959-1968, 2010.

Shahsavari, A., and Akbari, M., Potential of solar energy in developing countries for reducing energy-related emissions. *Renewable and Sustainable Energy Reviews*, vol. 90, pp.275-291, 2018.

Sivamadhavi, V., and Selvaraj, R.S., Prediction of monthly mean daily global solar radiation using Artificial Neural Network, *Journal of earth system science*, vol. 121, pp.1501-1510, 2012.

Wibawa, A.P., Utama, A.B.P., Elmunsyah, H., Pujianto, U., Dwiyanto, F.A. and Hernandez, L., Time-series analysis with smoothed Convolutional Neural Network, *Journal of big Data*, vol. 9, no. 1, p.44, 2022.

Yang, D., Wang, W., Gueymard, CA., Hong, T., Kleissl, J., Huang, J., Perez , MJ., Perez, R., Bright, JM., Xia, XA., and van Der Meer, D. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renewable and Sustainable Energy Reviews*, vol. 161, 2022.