Energy Proceedings
Vol 51, 2025
ISSN 2004-2965

Prediction of Daily Solar Irradiance for Solar Energy Approximation using LSTM Neural Network#

Roslin Mahmud Joy[1], Md. Faruk Abdullah Al Sohan[1], Afroza Nahar[1*], and Nusrat Jahan Anannya[1]

1 American International University-Bangladesh, Dhaka -1229, Bangladesh

(Corresponding Author: afroza@aiub.edu)

ABSTRACT

Solar energy systems suffer from the unstable nature of solar irradiance. This paper proposes a novel approach to predict solar irradiance using a sequential model with a Long Short-Term Memory (LSTM) network. The model uses the mean squared error (MSE) loss function and the Adam optimization algorithm to provide robustness to noise and efficient computation of gradients. The model aims to forecast solar irradiance for 2023 in Dhaka, Bangladesh, using weather data and sunlight duration data from NASA Power and Data. GISS NASA. The dataset includes daily records from 2000-2023. The proposed LSTM model achieves 5.47% more accuracy than RandomForestRegressor in terms of root mean squared error (RMSE) to predict 1 year of irradiance data, indicating a substantial degree of precision in forecasting the objective variable. The results corroborate previous research highlighting the advantages of utilizing temporal dependencies and historical data to achieve accurate solar energy forecasts.

Keywords: solar energy, deep learning, daily irradiance prediction, LSTM, Adam optimization.

NONMENCLATURE

Abbreviations	
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
LSTM	Long Short-Term Memory
RNNs	Recurrent Neural Networks

1. INTRODUCTION

Solar power generation, grid stability, and energy market operations depend on accurate solar irradiance predictions [1]. Bangladesh faces power crisis, especially in rural areas, wherein solar energy could be a promising solution due to abundant solar incidence. However, the scarcity of meteorological devices for direct solar irradiation measurement makes alternative estimation models crucial for solar project development [2]. Machine learning and deep learning techniques, specifically Long Short-Term Memory (LSTM) networks, showed significant potential in predicting time series data, such as solar irradiation [3].

Predicting the daily solar irradiance is very important for getting the most out of solar energy production. Lei Li et al. [4] proposed some methods for dividing training datasets and utilizing an LSTM-based model to forecast solar irradiance. They showed an error analysis to identify the most suitable strategy for splitting the dataset. Ghizlene Cheikh et al. [5] conducted a comparison of many deep learning models, such as LSTM, AE-LSTM, and CNN-LSTM autoencoder. They explained that the CNN-LSTM autoencoder exhibited the best level of accuracy, achieving an RMSE of around 93%. Swarna Venkata et al. [6] used LSTM to produce superior prediction rates with minimum error. It highlights the significance of precise solar radiation prediction for sustainable development and efficient energy usage. Jeon and Kim [7] proposed a hybrid model that combines LSTM and Bi-LSTM layers. This model demonstrated excellent outcomes even when trained with data from various geographies, obtaining an RMSE of 69.5 W/m². Firas Gerges et al. [9] developed a Bayesian deep learning framework called DeepSI. This developed framework combines bidirectional LSTM autoencoders with a transformer and Monte-Carlo dropout sampling. It is used to forecast daily solar irradiance. It would be effective for long-term forecasts until the year 2099 [8]. Veysel Gider et al. conducted research and another research at Dicle University used LSTM models to predict solar irradiance. These studies highlighted the significance of precise GSR estimations for planning power generation and ensuring system resilience.

This paper introduces a novel technique for predicting the daily solar radiation received in Dhaka, Bangladesh. The approach employs an LSTM-based model that uses historical meteorological data and the length of sunlight to forecast solar irradiance, aiming to

This is a paper for the 16th International Conference on Applied Energy (ICAE2024), Sep. 1-5, 2024, Niigata, Japan.

improve prediction accuracy and support the development of solar energy projects in the region.

2. MATERIALS AND METHOD

2.1 LSTM Neural Network Overview

We proposed a long short-term memory (LSTM) networks-based machine learning methods for Predicting daily solar irradiance. LSTMs are a specific category of recurrent neural networks (RNNs) that can learn from and generate forecasts using time series data. For solar irradiation forecasting, LSTMs specifically remember information over extended time periods. Solar irradiation data often exhibit seasonal patterns and long-term trends that are difficult to capture with conventional models.

2.2 Data Collection and Preprocessing

The data for this study was sourced from two primary databases: NASA Power [1] and GISS NASA [2]. NASA Power provided comprehensive weather data, including daily records of temperature, wind speed etc. GISS NASA supplied the daily sunshine duration data. The dataset spans from the year 2000 to 2023, encompassing daily records for Dhaka, Bangladesh, with coordinates 23.77°N latitude and 90.39°E longitude.

To create a unified dataset, the data from NASA Power and GISS NASA were merged using the common Date column. This merging process ensured that each record corresponded to the same day across both datasets, providing a comprehensive view of the weather and sunshine conditions. The merged dataset included approximately 8766 rows, each representing a day's worth of data.

We selected features vector x_d on a specific day d. Where x_d denoted by $[x_{d,1}, x_{d,2}, x_{d,3}, ..., x_{d,7}]$. Items of the feature vectors are as follows:

- $x_{d,1}$: timestamp of current date
- $x_{d,2}$: month of the day
- $x_{d,3}$: sunrise time of the day
- $x_{d,4}$: sunset time of the day
- $x_{d,5}$: the average temperature of the day
- $x_{d,6}$: the average wind speed of the day
- $x_{d,7}$: the average surface pressure of the day
- $x_{d,8}$: sunlight weighted Cosine of Zenith Angle

We predicted the daily solar irradiance value y_d based on the selected feature vector X_d. The selection of these parameters was significantly influenced by the correlation matrix shown in Fig. 1 which provided valuable insights into the relationships between different variables, allowing us to choose those with high correlation for our feature vector.

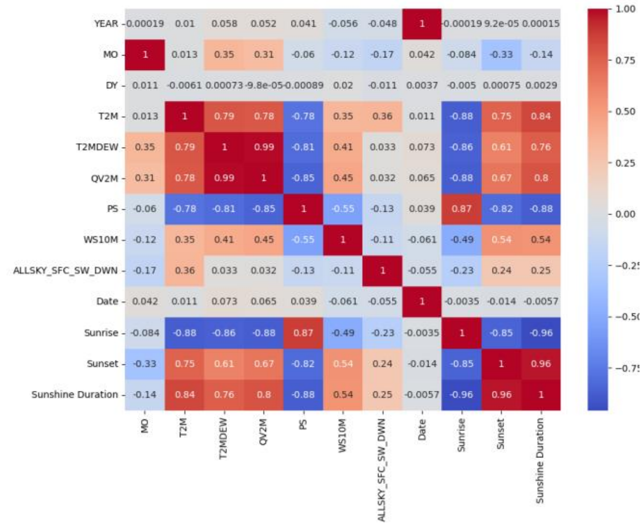

Fig. 1 Correlation matrix of feature vector

The month was specially extracted from the Date column to capture seasonal variations in solar irradiance. This comprehensive dataset was then split into training and testing sets, with data from the years 2000 to 2022 used for training and data from the year 2023 reserved for testing. This approach ensured that the model was trained on a robust dataset and could be evaluated on unseen data to assess its predictive performance.

2.3 Model Preparation and Training

Unlike traditional RNNs, LSTMs have a unique design that prevents the vanishing gradient problem, making them effective at capturing long-term dependencies in the data. In this study, we used the Keras deep learning library from Tensorflow to construct the Sequential model with an LSTM layer for implementing the prediction models.

The model is trained using the Mean Squared Error (MSE) loss function, which measures the average squared difference between the predicted and actual values. This loss function is suitable for regression problems like ours, where the goal is to minimize the difference between the predicted and actual values. The Adam optimization algorithm is used to update the model parameters. Adam combines the advantages of two other extensions of stochastic gradient descent: AdaGrad and RMSProp. It computes adaptive learning rates for different parameters, providing robustness to noise and efficient computation of gradients.

For comparison against traditional machine learning methods, in addition to the LSTM deep neural network, we also employ a RandomForestRegressor model for solar irradiance prediction. The RandomForestRegressor is a meta estimator that fits several classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

2.4 Experiment 1

In the first experiment, we predict using LSTM neural network, our primary model. This model is composed of sequential models with hidden neurons of 30. Input layers of this network have 7 features. Output of this network with one activation function had one neuron. The maximum number of epochs was set to 192, shown in Fig. 2.

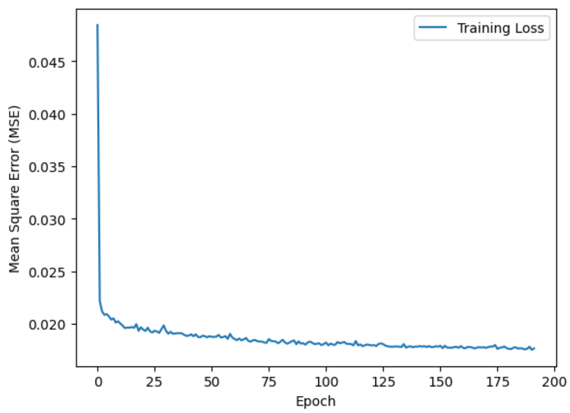

Fig. 2 Training loss graph of LSTM network

2.5 Experiment 2

In the second experiment, we utilized the RandomForestRegressor model for predicting solar irradiance. This model was configured with 100 estimators and a random state of 42 to ensure reproducibility. The training process involved fitting the model to the training dataset, which included the same features used in the LSTM model. After training, the model was used to make predictions on the test dataset. The RandomForestRegressor performance was then evaluated and compared to the LSTM model to assess its efficacy in solar irradiance prediction.

3. RESULTS

3.1 Model Performance

The performance of the proposed LSTM model was evaluated using the test dataset for the year 2023. The model's accuracy and robustness were assessed through various metrics, with a primary focus on the Root Mean Squared Error (RMSE).

The LSTM model demonstrated exceptional accuracy in predicting daily solar irradiance. The model achieved a Root Mean Squared Error (RMSE) of 43.48 w/m^2, indicating a minimal difference between the predicted and actual values. This low RMSE underscores the model's ability to accurately capture the temporal dependencies and patterns in the solar irradiance data, shown in Fig. 3.

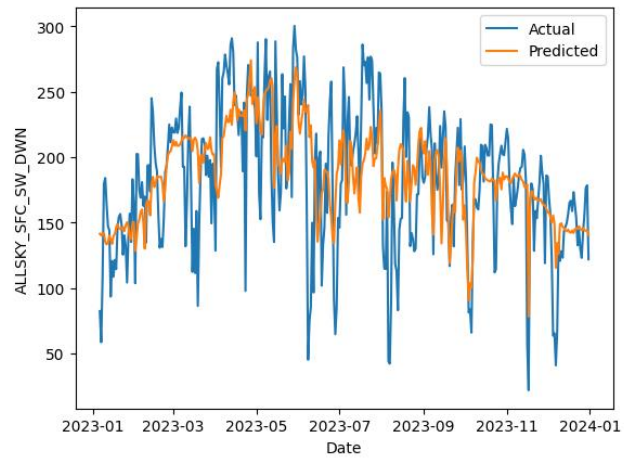

Fig. 3 Actual and Predicted value of Solar Irradiance

3.2 Comparison with traditional model

To benchmark the performance of the long short-term memory (LSTM) model, we compared it against traditional machine learning model, RandomForestRegressor, which achieved an RMSE of 46 W/m². The results are summarized in Table 1.

Table 1: Root Mean Squared Error value of models

Model	RMSE
LSTM	43.48 w/m^2
RandomForestRegressor	46 w/m^2

The LSTM model outperformed the traditional model, achieving the lowest RMSE. Specifically, the LSTM model was approximately 5.47% more accurate than the RandomForestRegressor model.

4. CONCLUSIONS

We have developed a robust method for forecasting daily solar irradiance by utilizing a LSTM neural network. Utilizing a large dataset of meteorological and solar irradiance information spanning from 2000 to 2022, our model achieved a high accuracy. Specifically, it achieved a RMSE of 43.48 W/m² for the year 2023 in Dhaka, Bangladesh. In comparison, the standard

RandomForestRegressor model only achieved an RMSE of 46 W/m².

This research highlights the substantial potential of advanced deep learning models to improve solar irradiance prediction accuracy, which is crucial for advancing solar energy projects, especially in areas lacking direct measurement infrastructure. Future studies could further enhance prediction accuracy by incorporating hybrid models and factors such as cloud cover and atmospheric conditions, aiding the global transition to sustainable energy.

ACKNOWLEDGEMENT

Authors would like to express their gratitude to American International University–Bangladesh for providing technical support to carry out this research.

REFERENCE

[1] Chowdhury, M. S., Rahman, K. S., Chowdhury, T., Nuthammachot, N., Techato, K., Akhtaruzzaman, M., Tiong, S. K., Sopian, K., & Amin, N. (2019). An overview of solar photovoltaic panels' end-of-life material recycling. *Energy Strategy Reviews*, 27, 100431.

[2] Rabbi, K. M., Nandi, I., Saleh, A. S., Faisal, F., & Mojumder, S. (2016). Prediction of solar irradiation in Bangladesh using artificial neural network (ANN) and data mapping using GIS technology. *4th International Conference on the Development in the in Renewable Energy Technology (ICDRET'2016)* (pp. 1-6). IEEE.

[3] Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802-810.

[4] Li, L., Zhang, L., Li, D., Zhang, W., & Yue, H. (2024, February). The influence of training datasets on LSTM-based irradiance prediction. In *Journal of Physics: Conference Series* (Vol. 2703, No. 1, p. 012011). IOP Publishing.

[5] Cheikh, G., Hadj, N. B., Ammar, B., Bentounes, K. A., Bentounes, H. A., & Larbi, A. (2023, October). Deep CNN-LSTM Autoencoder for Solar Irradiation Prediction. In *2023 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE)* (pp. 1-6). IEEE.

[6] Aditya, S. V. N., Bhuvaneswari, R., Natarajan, B., & Dhavakumar, P. (2023, August). Solar Irradiance Prediction Model Based on LSTM. In *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)* (pp. 1-5). IEEE.

[7] Jeon, B. K., & Kim, E. J. (2023). Performance Evaluation of LSTM Based Solar Irradiance Prediction Models without Long-term Local Weather Data. *Journal of the Korean Solar Energy Society*, *43*(5), 43-59.

[8] Gerges, F., Boufadel, M. C., Bou-Zeid, E., Nassif, H., & Wang, J. T. (2023, January). Deep Learning-Enabled Prediction of Daily Solar Irradiance from Simulated Climate Data. In *Proceedings of the 2023 7th International Conference on Machine Learning and Soft Computing* (pp. 102-109).

[9] Gider, V., Budak, C., Izci, D., & Ekinci, S. (2022, October). Daily solar radiation prediction using LSTM Neural Networks. In *2022 Global Energy Conference (GEC)* (pp. 168-172). IEEE.

[10] NASA POWER Project. (2024). Daily POWER Daily API v2.5.9 OAS3. Retrieved on June 12, from https://power.larc.nasa.gov/api/pages/?urls.primaryName=Daily

[11] Russell, G. (2024). Sunrise, Sunset, and Insolation Data. Retrieved on June 10, 2024, from https://data.giss.nasa.gov/modelE/ar5plots/srlocat.html