Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Enhancing Solar Irradiance Forecasting Using LSTM and Meteorological Data

Ezmin Abdullah[1,2], Mohammad Irham Hafiez Mohammad Ahza Wee[2], Roslina Mohamad[1,2], Nabil M. Hidayat*[,2]
[1] *Wireless High-Speed Network Research Interest Group (RIG), Universiti Teknologi MARA, 40000 Shah Alam, Selangor, Malaysia*
[2]*Faculty of Electrical Engineering, Universiti Teknologi MARA, 40000 Shah Alam, Selangor, Malaysia*
**Corresponding Author: mnabil@uitm.edu.my*

ARTICLE INFO

Received: 29 Dec 2024

Revised: 15 Feb 2025

Accepted: 24 Feb 2025

ABSTRACT

Introduction: Solar energy is an abundant and sustainable resource that plays a crucial role in the global shift towards renewable energy. However, solar irradiance, a key factor in solar power generation, fluctuates throughout the day, affecting the efficiency of solar energy systems. Accurate prediction of solar irradiance is essential for optimizing solar energy generation and ensuring grid stability.

Objectives: This study aims to improve the accuracy of solar irradiance forecasting using a Long Short-Term Memory (LSTM) model. The objective is to address the limitations of traditional forecasting models and explore the integration of various meteorological inputs for more precise predictions.

Methods: The LSTM model was developed using historical solar irradiance data and meteorological parameters, including temperature, humidity, and wind speed, sourced from NASA's POWER Solar database. Data preprocessing techniques like MinMaxScaler normalization were applied, and the model was trained using 70% of the data and tested on 30%. The LSTM network incorporated layers with 256, 128, and 64 units, optimized using techniques like EarlyStopping and ReduceLROnPlateau to avoid overfitting.

Results: The LSTM model demonstrated strong predictive performance, achieving an RMSE of 47.58, MAE of 22.64, and an R-square value of 96.94%. Compared to traditional Support Vector Regression (SVR), the LSTM model outperformed with a 28.79% improvement in RMSE and a 48.81% improvement in MAE. The model's ability to capture temporal dependencies and nonlinear interactions in solar irradiance data was confirmed through evaluation metrics.

Conclusions: The LSTM model successfully enhanced solar irradiance forecasting, providing more accurate predictions for renewable energy applications. Despite its promising results, future research can explore additional weather parameters and hybrid machine learning models to further improve accuracy and generalizability.

Keywords: solar irradiance, LSTM, SVR, Meteorological Data, Renewable Energy.

INTRODUCTION

In recent years, the smooth functioning of the Fourth Industrial Revolution (4IR) has remained challenging without a reliable energy supply to power the technologies and systems that drive it, despite AI being hailed as its new 'electricity [1]. While a significant portion of the world's electricity is still generated from non-renewable energy sources, the share of renewable energy is steadily increasing [2][3]. This shift is driven by the increasing demand for Renewable Energy Sources (RES) to mitigate the adverse environmental impacts of fossil fuels and other non-renewable resources. [4].

Among the various renewable energy sources, solar energy is the most common type of renewable energy since the energy source is abundant, and the costs of installation are relatively low [5]. Solar energy adoption has been growing rapidly across residential and industrial sectors, outpacing other renewable sources. However, its random (fluctuating radiation) character still threatens the power system and the electrical grid's capability to work efficiently [6]. This difficulty arises from the fact that PV panels performance in generating is a directly proportional to sun radiation [7]. Solar irradiance also known as Global Horizontal Irradiance (GHI) varies with time of the day. During the day their values range from zero to approximately 1,100 Wats/m2 with respect to area. At night it too depresses to zero. In predictive research one of the key factors for the stability of the electricity power generation is to predict the solar irradiance.

Past studies have employed various forecasting techniques such as Support Vector Regression (SVR). A study in paper [8] applied SVR to forecast solar irradiance with weather parameter inputs and achieving moderate success of 70 W/m^2 RMSE. However, the study reported difficulty in optimizing kernel parameters thus limiting its scalability for large datasets. Other study in paper [9] uses hybrid approaches combining machine learning and signal processing. A hybrid SVR-wavelet model that improved short-term forecasting accuracy is capable to achieve RMSE of 65 W/m^2. While hybrid models address data noise, they require significant computational resources and complex tuning processes. In additional to that, Artificial Neural Networks (ANNs) based model that is developed in paper [10] for irradiance prediction in urban areas is capable to achieve an RMSE of 75 W/m^2. Despite their utility, ANNs struggle to model sequential data effectively due to vanishing gradient issues thus making their suitability for time series limited. LSTM models have emerged as the leading choice for sequential data modeling. Study in [11] implemented LSTM model to forecast hourly solar irradiance and able to achieve an RMSE of 50 W/m^2 and outperforming both ANNs and SVR. Similarly, a study in paper [12] integrated meteorological variables such as wind speed and humidity into an LSTM model, improving its predictive accuracy and able to get an RMSE of 48 W/m^2. Although there are still issues with maximizing training effectiveness and generalizability, these studies show that LSTM can manage temporal dependencies and nonlinear interactions successfully.

Therefore, in this paper, proposed LSTM model addresses these limitations by integrating comprehensive meteorological inputs such as surface pressure, clear sky surface shortwave downward irradiance, temperature, wind speed and humidity to enhance prediction accuracy. In additional to that, techniques like EarlyStopping and ReduceLROnPlateau to prevent overfitting and optimize training processes efficiently. This study creates a strong framework for solar irradiance forecasting by filling in the gaps found in previous studies, which helps to optimize renewable energy sources and build climate resilience.

OBJECTIVES

In this paper, Long Short-Term Memory (LSTM) network, a variant of Recurrent Neural Network (RNN), was applied in forecasting solar irradiance. LSTM is a Deep learning (DL) network developed to handle issues with vanishing & exploding gradients common with a conventional RNN [13]. DL has been proved to be a very useful class of models for a number of challenging learning tasks. It has also been shown to be very effective at retrieving hierarchical information from the spatial-temporal data domain with tens of millions of parameters [8]. In this study, DL was utilized to anticipate hourly sun irradiance due of its great power as a technique for pattern/sequence identification challenges. The outcomes from the Support Vector Regression (SVR) method and the proposed LSTM model were compared. The root mean squared error (RMSE) and R-square findings from the various tests were taken into consideration as comparison performance evaluators to ascertain whether forecasting algorithm is more accurate in terms of error rate [14]. The dataset in this study is obtained from NASA POWER Project's Data Access Viewer (DAV).

METHODS

The first phase is data preparation. This phase involves collecting and preprocessing the historical weather data to ensure it's suitable for model training. Historical solar irradiance data and meteorological parameters like temperature, humidity, and wind speed are sourced from NASA POWER's DAV. The location collected is at the Universiti Teknologi MARA (UiTM) Shah Alam, at the latitude of 3.0719 and longitude 101.501. The data is normalized using the MinMaxScaler to ensure uniform scaling and the dataset is split into training (70%) and testing (30%) for model development and evaluation.

The second phase of the research study focuses on designing and training the LSTM model to predict solar irradiance. The LSTM model includes three layers with 256, 128, and 64 units, respectively. A dense output layer is used to generate the final prediction. The training configurations include the Adam optimizer, Huber loss function, 200 epochs and a batch size of 32. The Hyperparameter such as learning rate are fine-tuned using grid search. Furthermore, additional techniques were implemented to enhance the training process such as EarlyStopping and ReduceLROnPlateau. By using EarlyStopping, it can help fight overfitting by halting the training if no improvement

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

was observed after 20 epochs. On the other hand, ReduceLROnPlateau will reduce the learning rate by a factor of 0.5 if validation loss plateaued for 10 consecutive epochs.

The third phase of the research study is to evaluate the model's performance and validate its predictions using metrics and visualizations. The evaluation metrics used is R-square score, it measures the proportion of variance explained by the model. Next is root mean square error (RMSE). RMSE measures the average difference between values predicted by the model to the actual values while Mean Absolute Error (MAE) measures the average absolute difference between predictions and actual values.

A. LSTM Neural Network

The structure of the LSTM network model consists of gates with several different functions and contains information about previous states. Information in this network structure is written, stored, or read by a cell that acts as a memory as in Fig.1 [13]. The corresponding part in the cell enables it to decide whether to keep the information received when these gates are opened and closed in the read, write and erase state. For this purpose, it acts according to the signals received, blocks, and transmits the information according to its power, and carries out the import process by filtering its own quantity and weights. The input, hidden and output layers in the LSTM structure are shown in Fig. 2.

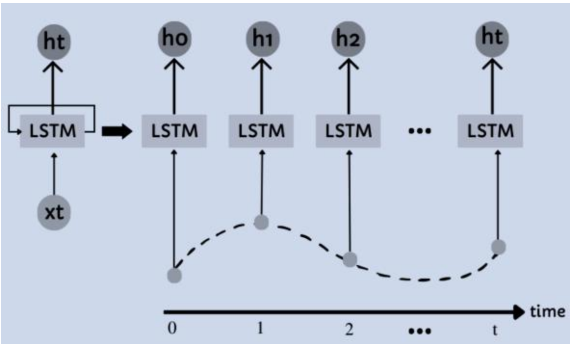

Fig. 1. LSTM Input-Output

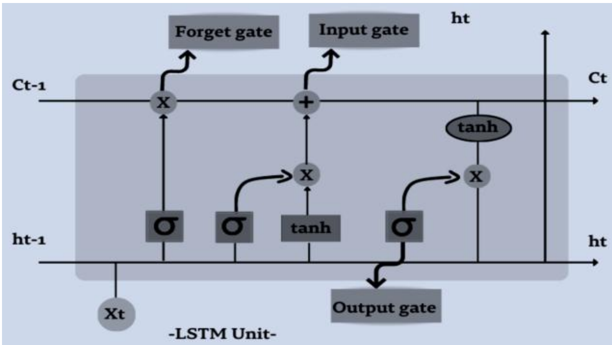

Fig. 2. LSTM hidden and output layers

B. Forget gate activation

The forget gate decided by logistic function what information to keep or discard on the state of cell ct. The mathematical expression of this gate is expressed as

$$f_t = \sigma\,(W_f\,[h_{t-1}, x_t] + b_f) \qquad (1)$$

where σ is the activation function, W_f is the weight of forget gate, b_f is the bias of forget gate, x_t is the input at time t, $ht{-}1$ is the hidden layer output at time t-1.

Examining the input gate shown in Fig. 3, the LSTM block structure decides whether to update the input values. To do this, i_t and C_t is calculated according to (2) and (3) respectively [14].

$$i_t = \sigma\,(W_i\,[h_{t-1}, x_t] + b_i) \qquad (2)$$

$$C_t = tanh\,(W_c\,[h_{t-1}, x_t] + b_c) \quad (3)$$

where, b_i is the bias of input gate, W_c is the weight of cell, b_c is the bias of cell.

Journal of Information Systems Engineering and Management
2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

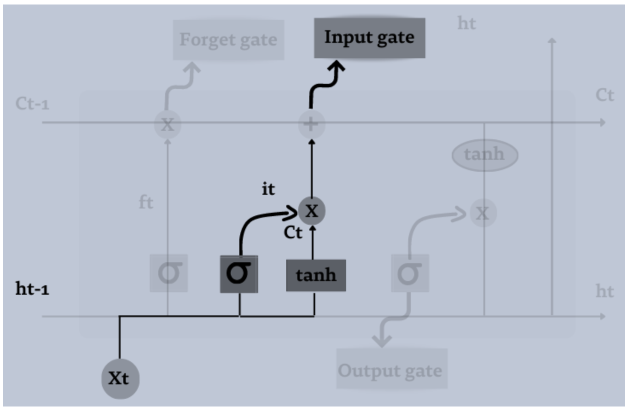

Fig. 3. LSTM Unit Input Structure

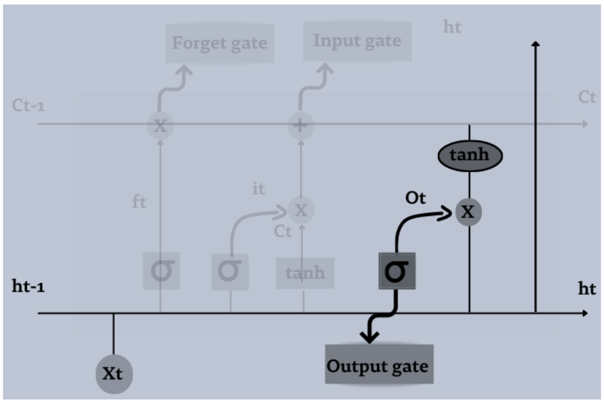

Fig. 4. LSTM Unit Output Structure

Fig. 4 shows the output structure with a tanh that decides which part of each cell state to select. The forget gate h_t and the output gate o_t are scaled with a logistic function. These operations are performed with (4) and (5) respectively [15, 16].

$$o_t = \sigma\ (W_o\ [h_{t-1}, x_t] + b_o) \qquad (4)$$

$$h_t = o_t + \tanh\ (C_t) \qquad (5)$$

where, Wo is the weight of output gate, bo is the bias of the output gate.

RESULTS

This section discusses the LSTM model's performances. The regression lines show how the datasets fit in the models. Performance measures including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and normalized RMSE were used in this investigation. Visual Studio Code was used for developing the models using Python. The forecasting results based on the 5 years daily solar irradiance data that is collected from the NASA POWER's DAV from 1 January 2019 - 31 december 2023.

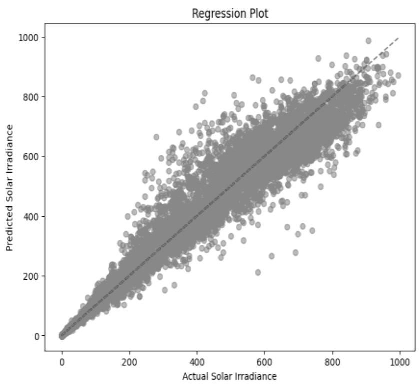

Fig. 5. LSTM regression plot

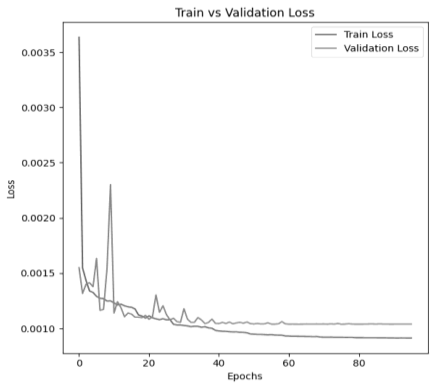

Fig. 6. Training vs validation loss plot

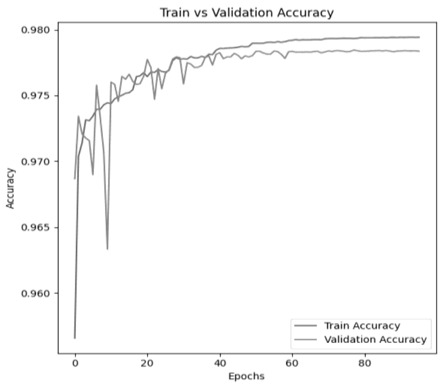

Fig. 7. Training vs validation accuracy plot

The regression plot in Fig. 5 shows actual versus predicted solar irradiance values, with points closely clustered around the regession line, indicating the model's accuracy in capturing the relationship between inputs and solar irradiance. The high R-square score of 0.9694 confirms that 96.94% of the data's variance is explained by the model. While a few points deviate from the line, the overall closeness suggests strong predictive performance.

Figure 6 illustrates the training and validation error across epochs, showing steadily decreasing loss values, indicating effective learning. The minimal gap between training and validation losses confirms the model's good generalization and low overfitting, consistent with an RMSE of 47.58. Figure 7 shows the model's predictions during training and validation, with both accuracies increasing over epochs, indicating improved performance with more training. The convergence of the accuracy curves at later epochs confirms a good balance between fitting the training data and generalizing to new data. The model's MAE of 22.64 further highlights its accuracy in prediction.

Table 1. Number of LSTM Layers

No. Layer	RMSE	MAE	R2
1	49.67	24.62	96.80
2	48.73	23.02	96.86
3	47.58	22.64	96.94
4	48.23	23.15	96.88

Table 2. Number of Dropout

Dropout	RMSE	MAE	R2
0.0	47.58	22.64	96.94
0.2	48.20	23.30	95.40
0.4	49.46	24.97	95.12

Table 3. LSTM vs SVR

Metric	LSTM	SVR	Difference
RMSE	47.58	66.80	LSTM is 28.79% better
MAE	22.64	50.68	LSTM is 48.81% better
R-square	96.94	89.05	LSTM is 7.89% better

The ability of an LSTM model to capture patterns in sequential data, such as solar irradiance, depends on the number of layers. Each additional layer helps the model learn more complex features. A single LSTM layer with 128 units can model basic sequential patterns but struggles with long-term relationships. Adding a second layer improves the model's capacity to capture deeper patterns, but the performance gain is marginal, suggesting that the model isn't fully utilizing its potential. A three-layer LSTM strikes the best balance, extracting low-level patterns in the first layer, intermediate relationships in the second, and refining them into high-level features in the third. However, adding a fourth layer introduces complexity without substantial improvement and may lead to overfitting. The full comparison is shown in Table 1.

Dropout, a regularization technique, helps reduce overfitting by randomly deactivating neurons during training, forcing the model to learn more resilient features. However, excessive dropout can cause underfitting. Testing different dropout rates showed that no dropout (0.0) yielded the best results, with an RMSE of 47.58, MAE of 22.64, and an R-square of 96.94 as shown in Table 2. Increasing dropout to 0.2 degraded performance, and a dropout rate of 0.4 led to underfitting, increasing RMSE to 49.46 and reducing R^2 to 95.12. Thus, dropout was unnecessary for this model.

When compared to Support Vector Regression (SVR) as in Table 3, the LSTM model outperformed in all metrics. LSTM achieved an RMSE of 47.58, 28.79% better than SVR's 66.80, and an MAE of 22.64, 48.55% better than SVR's 44. The R^2 score of 96.94 for LSTM was 7.89% higher than SVR's 89.05, confirming that LSTM excels at handling the time-dependent and nonlinear nature of solar irradiance, making it ideal for renewable energy forecasting.

ACKNOWLEDGEMENT

This work was supported in part by Universiti Teknologi MARA.

DISCUSSION

The LSTM model successfully improved solar irradiance forecasting, offering accurate predictions for renewable energy applications with high R-square and low RMSE values. Despite its effectiveness, there is potential for further refinement by integrating more meteorological data and exploring hybrid machine learning models. These enhancements could further optimize prediction accuracy and ensure better generalization, supporting the development of more reliable solar energy systems.

REFRENCES

[1] A. C. Şerban and M. D. Lytras, "Artificial Intelligence for Smart Renewable Energy Sector in Europe—Smart Energy Infrastructures for Next Generation Smart Cities," in IEEE Access, vol. 8, pp. 77364-77377, 2020, doi: 10.1109/ACCESS.2020.2990123.

2025, 10(37s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

[2] A. Zoungrana and M. Çakmakci, “From non-renewable energy to renewable by harvesting salinity gradient power by reverse electrodialysis: A review,” in International Journal of Energy Research, 45(3), pp.3495-3522, 2021.
[3] Umair M, Hidayat NM, Sukri Ahmad A, Nik Ali NH, Mawardi MIM, et al. (2024) A renewable approach to electric vehicle charging through solar energy storage. PLOS ONE 19(2): e0297376. https://doi.org/10.1371/journal.pone.0297376.
[4] U.K. Pata, “Renewable and non-renewable energy consumption, economic complexity, CO2 emissions, and ecological footprint in the USA: testing the EKC hypothesis with a structural break.” In Environmental Science and Pollution Research, 28(1), pp.846-861, 2021.
[5] C. N. Obiora, A. N. Hasan, A. Ali and N. Alajarmeh, "Forecasting Hourly Solar Radiation Using Artificial Intelligence Techniques," in IEEE Canadian Journal of Electrical and Computer Engineering, vol. 44, no. 4, pp. 497-508, Fall 2021, DOI: 10.1109/ICJECE.2021.3093369.
[6] C. N. Obiora, A. Ali and A. N. Hasan, "Estimation of Hourly Global Solar Radiation Using Deep Learning Algorithms," 2020 11th International Renewable Energy Congress (IREC), 2020, pp. 1-6, DOI: 10.1109/IREC48820.2020.9310381.
[7] C. N. Obiora and A. Ali, "Effective Implementation of Convolutional Long Short-Term Memory (ConvLSTM) Network in Forecasting Solar Irradiance," IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, 2021, pp. 1-5, DOI: 10.1109/IECON48115.2021.9589934.
[8] R. Zhang, M. Feng, W. Zhang, S. Lu, and F. Wang, “Forecast of solar energy production-A deep learning approach,” presented at the 2018 IEEE International Conference on Big Knowledge (ICBK), 2018, pp. 73–82.
[9] H. Ali, A. Rahman, and M. Iqbal, “A hybrid SVR-wavelet model for short-term solar irradiance forecasting,” Renewable Energy, vol. 169, pp. 120–133, May 2021
[10] S. Sharma and A. Gupta, “ANN-based solar radiation forecasting for urban regions,” Renewable and Sustainable Energy Reviews, vol. 134, p. 110123, Jan. 2020.
[11] Z. Wu, Q. Li, and X. Xia, Multi-timescale forecast of solar irradiance based on multi-task learning and echo state network approaches. IEEE Transactions on Industrial Informatics, 2020 April 13.
[12] Liu, C. H., Gu, J. C., & Yang, M. T. (2021). A simplified LSTM neural networks for one day-ahead solar power forecasting. IEEE Access, 9,17174-17195.
[13] W. Di, A. Bhardwaj, and J. Wei, Deep Learning Essentials: Your hands-on guide to the fundamentals of deep learning and neural network modeling. Packt Publishing, 2018.
[14] Naga Aditya, S.V. et al. (2023) ‘Solar irradiance prediction model based on LSTM’, 2023 3rd Asian Conference on Innovation in Technology (ASIANCON) [Preprint]. doi:10.1109/asiancon58793.2023.10270057.
[15] Liu, C. H., Gu, J. C., & Yang, M. T. (2021). A simplified LSTM neural networks for one day-ahead solar power forecasting. IEEE Access, 9, 17174-17195.
[16] A. Demir, L. F. Gutiérrez, A. S. Namin and S. Bayne, "Solar Irradiance Prediction Using Transformer-based Machine Learning Models," 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 2022, pp. 2833-2840, doi: 10.1109/BigData55660.2022.10020615.