International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Research Article)

The Comparison of GRU and LSTM in Solar Power Generation Forecasting Application

An Duy Huynh * and Thanh Khuong Nguyen

Department of Mechatronics and Automation Engineering, Faculty of Electrical & Electronics Engineering, Ly Tu Trong College, Ho Chi Minh City, Viet Nam.

International Journal of Science and Research Archive, 2024, 13(01), 1360–1370

Publication history: Received on 10 August 2024; revised on 26 September 2024; accepted on 28 September 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.13.1.1831

Abstract

Precise forecasting of solar power generation is essential for optimizing the incorporation of renewable energy into the electrical grid and maintaining the stability of energy systems. This paper offers a thorough comparative examination of Long Short-Term Memory (LSTM) networks and Gated Recurrent Unit (GRU) networks in forecasting solar power generation. Both models were assessed for predicted accuracy and computational efficiency, employing essential performance indicators including Mean Absolute Error (MAE) and R-squared (R^2). The efficacy of LSTM and GRU in forecasting is assessed using a real-world dataset. Comprehensive experimental data demonstrate that the GRU model surpasses the LSTM model.

Keywords: Solar power forecasting; LSTM; GRU; Deep learning; Renewable energy

1. Introduction

Solar power generation forecasting is a crucial task for optimizing the integration of renewable energy sources into the electrical grid. Accurate forecasting helps in balancing supply and demand, reducing operational costs, and enhancing the stability and reliability of the power system. In recent years, advanced deep learning models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) have been widely adopted for time-series forecasting tasks, demonstrating significant improvements over traditional statistical methods.

LSTM and GRU, both derived from Recurrent Neural Networks (RNNs), are designed to capture long-term dependencies in sequential data while mitigating issues like the vanishing gradient problem (1, 4). LSTM models use a complex gating mechanism to control the flow of information, allowing them to effectively model temporal sequences (4). On the other hand, GRU simplifies the LSTM architecture by combining the forget and input gates into a single update gate, resulting in a more computationally efficient model without compromising performance (1).

Recent studies have extensively compared the performance of LSTM and GRU in various forecasting applications, including solar power generation. For example, Li et al. (7) compared LSTM and GRU for short-term photovoltaic power forecasting, concluding that while LSTM provides slightly better accuracy, GRU is faster to train and requires fewer computational resources, making it more suitable for real-time applications. Similarly, Feng et al. (2) explored hybrid models combining LSTM or GRU with CNN and found that these architectures significantly improved forecasting accuracy under different weather conditions.

Hybrid approaches have gained attention due to their ability to leverage the strengths of multiple models. Koprinska et al. (6) demonstrated that integrating LSTM with statistical methods in an ensemble model outperformed single-model approaches in solar power forecasting. Moreover, they highlighted that the choice between LSTM and GRU should be based on the specific application requirements, such as computational constraints and data characteristics. Other

* Corresponding author: An Duy Huynh

studies, such as Gigoni et al. (3), have reinforced these findings, showing that both LSTM and GRU can achieve high accuracy when combined with other machine learning techniques, such as support vector regression (SVR) and random forest.

Despite the advancements, there remains a lack of consensus on which model is superior for solar power forecasting. Some researchers argue that LSTM's ability to retain long-term dependencies makes it more suitable for capturing complex temporal patterns in solar power data (8). In contrast, others prefer GRU due to its simpler structure and faster convergence, which are advantageous when dealing with large datasets or limited computational resources (5).

This study aims to provide a comprehensive comparison of LSTM and GRU models in the context of solar power generation forecasting. By analyzing their performance across multiple datasets and under various conditions, this research seeks to offer insights into their relative strengths and weaknesses, thereby guiding the selection of appropriate models for different forecasting scenarios.

The main contributions of this research are as follows:

- Guidance for Model Selection: By providing a thorough comparison and analysis, this study serves as a guide for researchers and practitioners in selecting the most appropriate deep learning architecture for solar power forecasting, based on the specific needs of their applications, such as the trade-off between accuracy and computational efficiency.
- Foundation for Future Research: The findings of this research lay the groundwork for future studies to explore hybrid models, integrate additional environmental and operational factors, or develop more advanced neural network architectures to further enhance the accuracy and efficiency of solar power forecasting.
- Improved Predictive Metrics: The paper demonstrates the effectiveness of GRU models by achieving a lower Mean Absolute Error (MAE) and higher R-squared (R^2) compared to LSTM models. These improved metrics highlight the potential of GRU models for achieving high accuracy in solar power forecasting while reducing computational demands.

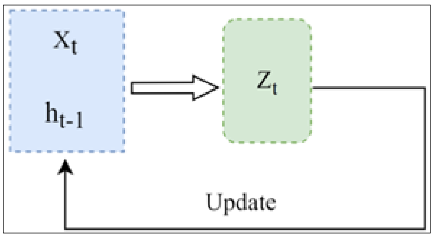


Figure 1 The architecture of standard recurrent neural network

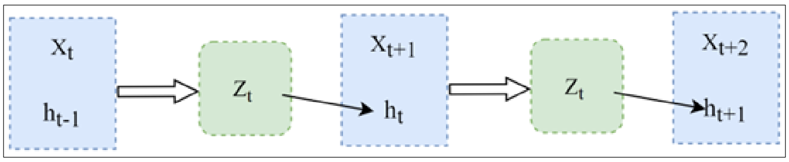


Figure 2 The feedforward structure of standard recurrent neural network

2. Relevant Theory

2.1. Recurrent Neural Network (RNN)

One sample at a time, a recurrent neural network (RNN) (9, 10) is a kind of neural network that can process sequential input. Because of this, RNNs are especially well-suited to capture the dynamic temporal patterns present in sequential data at various scales. In Figure 1, the hidden state from the previous time step and inputs from the current sample are processed by node. The RNN can learn dependencies over time by maintaining information across lengthy sequences thanks to this feedback loop.

Given an input sequence, the RNN produces a sequence of hidden states defined by:

$$h_t = \psi(z_t) = \psi(W_h h_{t-1} + W_x x_t + b) \qquad (1)$$

Where and are weight matrices, b is the bias vector, and the is a parameter of the RNN. A common choice for the activation function $\psi(\cdot)$ is the hyperbolic tangent function (tanh).

The RNN can be viewed as a collection of identical networks that communicate information to one another, as shown in Figure 2. RNNs are intended for sequential data modeling; yet, because of the well-known problems with vanishing and exploding gradients, training them with stochastic gradient descent (SGD) can be difficult. Gradient clipping is a common solution for the expanding gradient problem; however, it is less effective for the disappearing gradient problem. LSTM and GRU are two more sophisticated RNN variations that can be used to solve this problem (11).

2.2. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks are a type of Recurrent Neural Network (RNN) designed to handle the limitations of traditional RNNs, particularly in learning long-term dependencies in sequential data. By using an advanced gating mechanism, LSTMs are able to maintain and update information across lengthy sequences, in contrast to regular RNNs, which may experience issues with vanishing and exploding gradients. What follows focuses on the vanilla LSTM architecture with recurrent transition (12), which is provided by

$$\begin{bmatrix} z_t^o \\ z_t^i \\ z_t \\ z_t^f \end{bmatrix} = W_h h_{t-1} + W_x x_t + b \qquad (2)$$

$$c_t = \sigma(z_t^f) \odot c_{t-1} + \sigma(z_t^i) \odot \tanh(z_t) \qquad (3)$$

$$h_t = \sigma(z_t^o) \odot \tanh(c_t) \qquad (4)$$

Where and the initial states are the parameters of the network. The operator denotes the Hadamard product (element-wise multiplication). The is the sigmoid function

The vanilla LSTM differs from a conventional RNN in that it includes a second memory component, called c_t, which updates almost linearly. Because of its linearity, gradients flow more smoothly over time, which makes backpropagation easier to handle. Moreover, the LSTM updates its cell state via three different gates, in contrast to the RNN, which does so at each time step:

- The forget gate $\sigma(z_t^f)$ determines the amount of information from the previous time step that should be retained.
- The input gate $\sigma(z_t^i)$ regulates the flow of new information from the current input into the cell.
- The output gate $\sigma(z_t^o)$ controls which part of the cell state should be output as the next hidden state.

Because of its thoughtful design, the vanilla LSTM can reliably add or remove data over an extended length of time.

2.3. Gated Recurrent Unit (GRU)

The goal of Gated Recurrent Unit (GRU) networks is to reduce the computational complexity of Long Short-Term Memory (LSTM) networks while addressing the same difficulties of capturing long-term dependencies in sequential data. GRUs, like LSTMs, use gating methods to control information flow, which enables them to maintain, update, and output data in a selected manner over time. Nevertheless, GRUs are more computationally efficient while maintaining the capacity to handle long-term dependencies since they combine the functions of the input and forget gates into a single gate and do not have a separate memory cell.

The GRU architecture is built around two key gates:

- Reset Gate (r_t): The reset gate determines how much of the past information should be forgotten. When the reset gate is close to zero, the network forgets the previously computed hidden state, focusing only on the current input.

$$r_t = \sigma\left(W_r \cdot [h_{t-1}, x_t] + b_r\right) \quad (5)$$

- Update Gate (z_t): The update gate is responsible for deciding how much of the past information (from the previous hidden state) needs to be passed along to the future. It controls the balance between retaining the previous hidden state and incorporating the new candidate state.

$$z_t = \sigma\left(W_z \cdot [h_{t-1}, x_t] + b_z\right) \quad (6)$$

The detail of symbols is similar to LSTM.

3. Proposed method

The suggested method seeks to evaluate the efficacy of Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) networks in predicting solar power generation utilizing diverse environmental and meteorological input variables. The system's architecture, illustrated in the picture above, comprises three primary components: input signal capture, feature extraction, and forecasting through deep learning models (Fig.3).

The system acquires real-time data from numerous sensors that assess various environmental elements pertinent to solar power generation. The input characteristics comprise:

- Wind Speed: Assessed with an anemometer to evaluate the impact of wind on photovoltaic (PV) efficiency.
- Solar Radiation: Measured with a pyranometer to assess sunlight intensity, an essential element in electricity generation.
- Temperature and Humidity: Assessed by temperature and humidity sensors, as these factors affect the performance and efficiency of solar panels.

These inputs provide a comprehensive representation of the environmental variables affecting solar power generation, enabling the model to yield more accurate predictions. Each input data point is a time-series sequence, with length determined by the slicing window size.

The gathered input data is preprocessed and input into the feature extraction module. This module implements a sequence of transformations and normalizations on the raw data to improve feature representation. The retrieved attributes serve as inputs for the forecasting algorithms. The feature extraction procedure entails:

- Normalization: Adjusting the input data to an appropriate range for effective model training.
- Dimensionality Reduction: Minimizing the complexity of input features to avert overfitting and enhance model generalization.
- The core of the proposed method is the comparison between the GRU and LSTM networks in forecasting solar power generation. Both models are evaluated on the same dataset with identical feature sets to ensure a fair comparison.

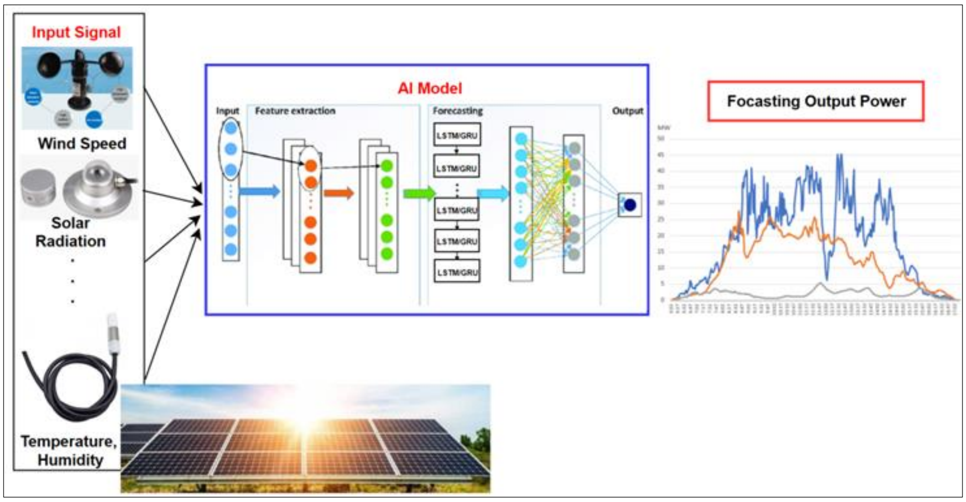


Figure 3 Block diagram of the model Forecasting

4. Evaluation

Both models are trained and tested using historical data to predict the future power output. The performance of the models is evaluated based on 2 key metrics: Mean Absolute Error (MAE), and R-squared (R^2) to determine which architecture is more suitable for this application.

4.1. Mean Absolute Error (MAE)

Mean Absolute Error (MAE) is a prevalent regression metric that quantifies the average amount of discrepancies between predicted and actual values, disregarding their direction. It is determined by the mean of the absolute discrepancies between the predicted values ($\hat{y}_i$) and the actual values (y_i).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \qquad (7)$$

Where n is the number of sample data.

4.2. R-squared (R^2)

R-squared (R^2) is a statistical metric that assesses the degree to which anticipated values align with actual data. It denotes the ratio of the variation in the dependent variable that may be anticipated from the independent variables. An R^2 score of 1 signifies that the model entirely accounts for the variance, whereas a value of 0 denotes a lack of explanatory capability.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \qquad (8)$$

Where:

$SS_{res} = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual Sum of Squares.

$SS_{tot} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ is the total Sum of Squares.

$\bar{y}$ is the mean of the actual values.

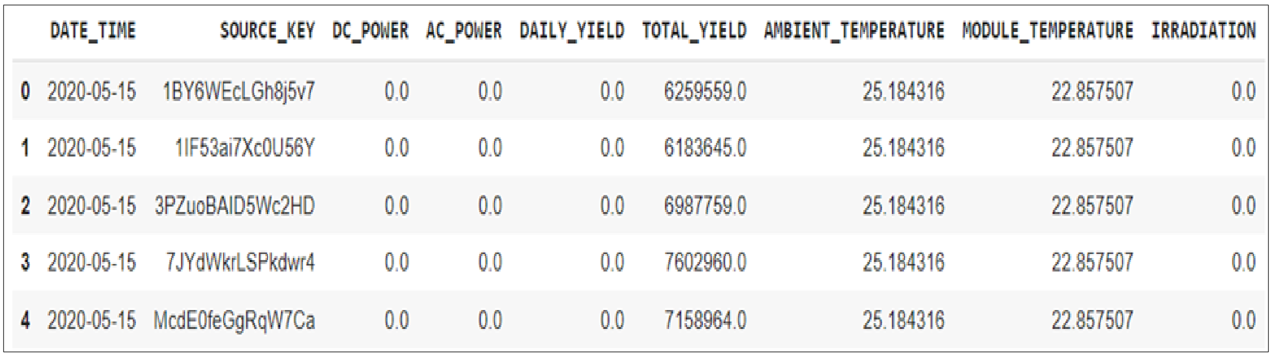
5. Experimental results

5.1. Dataset

The dataset contains solar power generation data collected from a single plant at 15-minute intervals over a period of 34 days (source: Energy_enthusiast, Solar Power Generation Data, Kaggle). The data includes various parameters that influence solar power generation, such as environmental conditions and operational metrics of the plant. The detail of key attributes in the dataset (Fig.4) are:

- DAILY_YIELD: The cumulative energy generated by the solar plant each day, measured in kilowatt-hours (kWh).
- TOTAL_YIELD: The total energy generated by the solar plant since the start of the data collection, measured in kilowatt-hours (kWh).
- AMBIENT_TEMPERATURE: The temperature of the environment surrounding the solar plant in degrees Celsius.
- MODULE_TEMPERATURE: The temperature of the solar panel modules in degrees Celsius.
- IRRADIATION: The amount of solar radiation received per unit area of the solar panels, measured in watts per square meter (W/m^2).

The correlation heatmap (Fig.5) visualizes the relationships between various features in the solar power generation dataset. Each cell in the heatmap represents the correlation coefficient between two features, with values ranging from 0 to 1. The heatmap confirms that IRRADIATION and MODULE_TEMPERATURE are crucial factors influencing the DC_POWER and AC_POWER outputs. The strong correlations between these variables suggest that they should be key inputs for any predictive models aimed at forecasting solar power generation. Conversely, variables like SENSOR_NUM and MINUTES have negligible correlations, indicating they might not be useful features for prediction tasks.



	DATE_TIME	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD	AMBIENT_TEMPERATURE	MODULE_TEMPERATURE	IRRADIATION
0	2020-05-15	1BY6WEcLGh8j5v7	0.0	0.0	0.0	6259559.0	25.184316	22.857507	0.0
1	2020-05-15	1IF53ai7Xc0U56Y	0.0	0.0	0.0	6183645.0	25.184316	22.857507	0.0
2	2020-05-15	3PZuoBAID5Wc2HD	0.0	0.0	0.0	6987759.0	25.184316	22.857507	0.0
3	2020-05-15	7JYdWkrLSPkdwr4	0.0	0.0	0.0	7602960.0	25.184316	22.857507	0.0
4	2020-05-15	McdE0feGgRqW7Ca	0.0	0.0	0.0	7158964.0	25.184316	22.857507	0.0

Figure 4 Structure of collected dataset

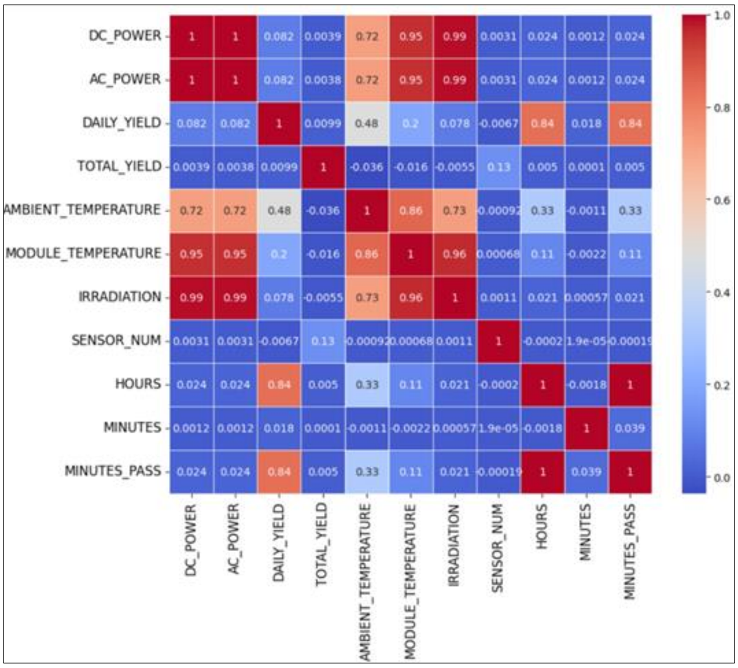


Figure 5 Correlation Heatmap

5.2. LSTM Prediction

Upon selecting a window size of 5, we segment the dataset using inputs of ambient temperature, module temperature, and irradiation, with the output being the total energy generated by the solar plant. We employ LSTM with an architecture similar to that depicted in the Fig.6.

In training and validation process (Fig.7, Fig.8), the training predictions and validation predictions closely align with the actual values across the entire range of time steps. This alignment suggests that the model generalizes well to unseen data during the training process. The smooth curve and consistent prediction accuracy imply that the model is well-regularized and not overfitting to the training data.

In testing result (Fig.9), the test predictions follow the actual values very closely, indicating a high level of accuracy on the test data. The MAE of 1.9642 indicates that, on average, the model's predictions deviate from the actual values by approximately 1.96 units of daily yield. This is a relatively small error, indicating a good predictive performance. The R^2 value of 0.9733 suggests that the model is highly effective in explaining the variability of the actual values. A value close to 1 indicates a very strong fit, meaning the model predictions are almost perfectly aligned with the actual values.

Layer (type)	Output Shape	Param #
lstm_2 (LSTM)	(None, 5, 128)	66,560
dropout_2 (Dropout)	(None, 5, 128)	0
lstm_3 (LSTM)	(None, 64)	49,408
dropout_3 (Dropout)	(None, 64)	0
dense_2 (Dense)	(None, 32)	2,080
dense_3 (Dense)	(None, 1)	33

Figure 6 Summary architecture of LSTM network

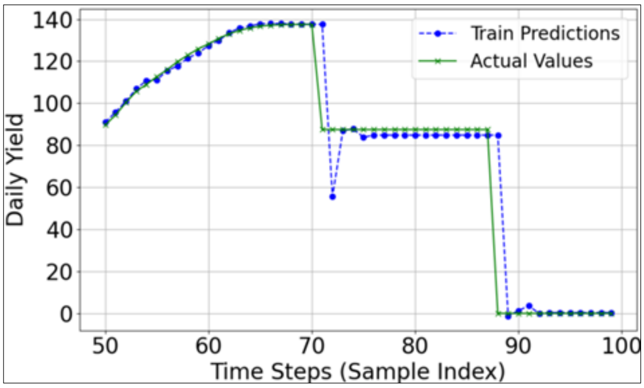


Figure 7 Result of training in LSTM

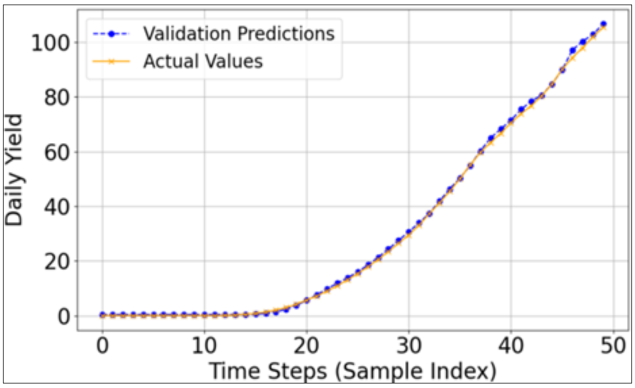


Figure 8 Result of validation in LSTM

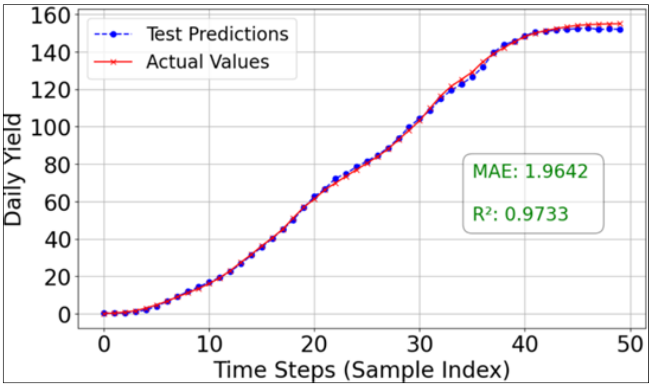


Figure 9 Result of testing in LSTM

5.3. GRU Prediction

For pair comparison, the GRU network will employ a comparable architecture to that of the LSTM (Fig. 10) and utilize the same dataset.

In training process and validation prediction (Fig.11, Fig.12) align well with the actual values for the majority of the data set, indicating that the model has learned the training data well. The consistent climbing trajectory and accurate forecast suggest that the model is finely calibrated and adept at managing the rising trend in daily yield efficiently.

The Mean Absolute Error (MAE) of 1.6180 signifies a marginal enhancement compared to the prior result of 1.9642 (Fig.13). A reduced MAE indicates a lesser average divergence from the actual data, implying enhanced accuracy.

The R-squared (R^2) value of 0.9836 marginally exceeds the prior result of 0.9733, signifying that the model accounts for about 98.36% of the variation in the test data. This indicates an even more robust alignment than previously.

Obviously, the improved MAE and R^2 values on the test set suggest that the current model (GRU) configuration has enhanced the predictive performance compared to the previous model (LSTM). This makes the current model more reliable for forecasting solar power generation.

Layer (type)	Output Shape	Param #
gru_2 (GRU)	(None, 5, 128)	50,304
dropout_6 (Dropout)	(None, 5, 128)	0
gru_3 (GRU)	(None, 64)	37,248
dropout_7 (Dropout)	(None, 64)	0
dense_6 (Dense)	(None, 32)	2,080
dense_7 (Dense)	(None, 1)	33

Figure 10 Summary architecture of GRU network

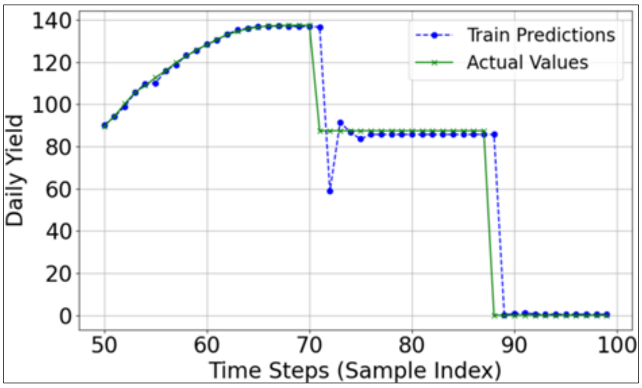


Figure 11 Result of training in GRU

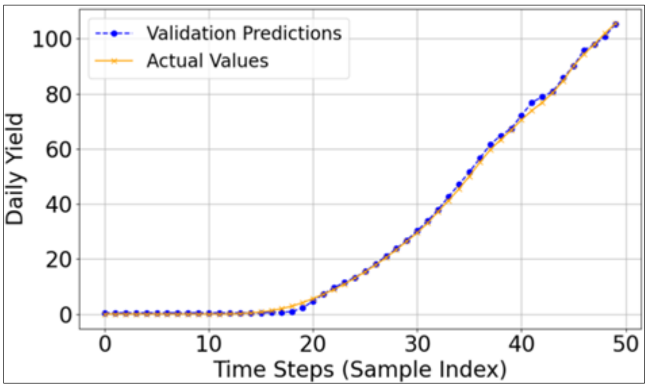


Figure 12 Result of validation in GRU

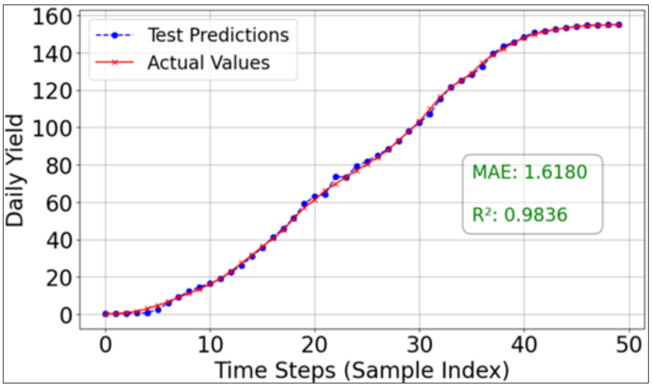


Figure 13 Result of testing in GRU

6. Conclusion

This study presents a comprehensive comparison between Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks for solar power generation forecasting. The results demonstrate that both LSTM and GRU models are highly effective in capturing the temporal patterns of solar power generation. However, the GRU model outperformed LSTM in terms of predictive accuracy and computational efficiency.

The experimental results show that the GRU model achieved a lower Mean Absolute Error (MAE) of 1.6180 compared to the LSTM's MAE of 1.9642, indicating a smaller average deviation from actual values. Furthermore, the R-squared (R^2) value of the GRU model was higher at 0.9836, suggesting that it explains more variability in the data compared to the LSTM's R^2 of 0.9733.

These findings suggest that while LSTM models may offer slightly better accuracy for certain complex patterns, the simpler architecture and faster convergence of GRU models make them more suitable for real-time applications and scenarios with limited computational resources. Therefore, the choice between LSTM and GRU should be guided by the specific requirements of the application, such as the desired balance between model accuracy and computational efficiency.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

Statement of informed consent

Informed consent was obtained from all individual participants included in the study.

References

[1] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[2] Feng, Y., Zhang, W., Liu, Y., & Tan, J. (2023). A hybrid model of CNN and LSTM autoencoder-based short-term PV power generation forecasting. Electrical Engineering. https://doi.org/10.1007/s00202-023-01472-0

[3] Gigoni, L., Testa, A., & Zizzo, G. (2023). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. Emerald Insight. https://doi.org/10.1108/IJESM-03-2022-0007

[4] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[5] Kim, J., Lee, D., & Park, J. (2023). Comparative study of LSTM and GRU for predicting solar irradiance. Renewable Energy, 204, 222-231. https://doi.org/10.1016/j.renene.2023.04.101

[6] Koprinska, I., Wu, D., & Wang, Z. (2023). Ensemble learning for solar power prediction: A comprehensive review and new perspectives. Renewable and Sustainable Energy Reviews, 142, 110849. https://doi.org/10.1016/j.rser.2021.110849

[7] Li, Y., Zhao, H., & Wang, Z. (2023). LSTM vs GRU for solar power forecasting: A comprehensive analysis on the effect of data pre-processing and feature engineering. Renewable Energy, 203, 1196-1208. https://doi.org/10.1016/j.renene.2023.02.015

[8] Yildirim, I., Dogru, A., & Yildirim, A. (2022). Solar power forecasting using LSTM and GRU models with Bayesian optimization. Energy, 238, 121902. https://doi.org/10.1016/j.energy.2021.121902

[9] I.Goodfellow, Y. Bengio, A. Courville, Deep learning, MIT press, 2016.

[10] S. Xiao, J. Yan, X. Yang, H. Zha, S. M. Chu, Modeling the intensity function of point process via recurrent neural networks., in: AAAI, 2017, pp. 1597–1603

[11] J. Schmidhuber, Deep learning in neural networks: An overview, Neural networks 61 (2015) 85–117.

[12] Zhao, Haitao, Shaoyuan Sun, and Bo Jin. "Sequential fault diagnosis based on LSTM neural network." IEEE Access 6 (2018): 12929-12939