Evaluating an Effectiveness of a Solar Power Plant Output Forecasting Model Based on LSTM Method Using Validation in Different Seasons of a Year in Vietnam

Duy Linh Bui[1], Quang Ninh Nguyen[1,2,*], Van Binh Doan[1,2], Tuan Khanh Pham[3], and Dinh Duong Le[4]

ARTICLE INFO

Article history:
Received: 18 August 2022
Revised: 22 November 2022
Accepted: 26 December 2022

Keywords:
LSTM
Forecasting
Validation
Solar power plant
P/GHI factor

ABSTRACT

This paper evaluates the effectiveness of the Long Short-term Memory (LSTM) method using the P/GHI (power/Global Horizontal Irradiance) factor and validation in the training process to forecast the generating capacity of a solar power plant (SPP) in case of changing the weather by seasons and changing geographic position conditions in Vietnam. The parameter matrix of the LSTM model for a SPP in the central part of Vietnam, where the climate is different from that of the southern part of Vietnam, is built based on the training and data filtering methods using the P/GHI factor and validation. The obtained model is applied to forecast the generating capacity of this SPP at specific climatic times of the region: sunny season, and rainy season, one day and one month, correspond. The input parameters are the real weather parameters and real output power in the past of the plant. Forecast results are compared with real output power data of the plant in the past. The calculation results show that the predicting method is still highly effective with MAPE error in the case of forecasting the generating capacity of SPP s in the period from 6:00 to 18:00 is about 8.0% to 9.4%. Comments and orientations for further research to be able to apply forecasting software for solar power plants in actual operation are proposed.

1. INTRODUCTION

The total capacity of renewable power sources in the world as of December 2021 reaches 3068 GW [1] (Fig.1), in which solar energy accounted about 855 GW. Of the 260GW total renewable power capacity added by 2021 in the world, Asia contributes 60%.

Vietnam is one of the countries in Asia focusing on encouraging the exploitation and development of solar energy sources. With preferential policies on FIT prices over the years, according to statistics from EVN (Vietnam Electricity), in 2019 there were 5052 MW, 105 projects of SPPs integrated into the national power grid. The capacity of these SPPs is about 30-50 MW and located in areas having high solar radiation such as the Central and Southern. Up to now, according to the statistics of the National Load Dispatch Centre, A0, there are 146 SPPs in Vietnam with a total capacity of about 8800 MW.

Depending on the needs of renewable energy (RE) plants and how the electricity market is managed in each country, owners of renewable power plants can choose to participate in the energy market or the reserve market [2]. Regardless of market participation, RE plants that do not use battery storage systems need a reasonable adjustment strategy to achieve the highest economic efficiency, avoiding the reduction of generating capacity [3]. In order to have a good competitive bidding strategy when entering the market, it is necessary to accurately forecast the generating capacity as well as the load demand [4]-[6].

There are many forecasting methods, the most popular today are persistence, physical and statistical models [2]. The persistence model is the simplest when using only an existing sample from the past [7]. In the Physical method, the future meteorological parameters are forecasted, through which the predicted output power is determined through the formulas [8, 9]. Statistical methods determine the correlation between past and present data to predict the output power. Statistical methods have two main techniques: time series [10]-[12], and Artificial Intelligence [13]-[18]. Which, the second group of techniques is being used by many authors [19]. Authors in [20] proposed a solution to improve the training process of the LSTM network to predict the generating capacity of industrial-scale solar power plants in Vietnam. The authors used the P/GHI coefficient to process input data as well as validation to increase accuracy and reduce model training time.

[1]*Faculty of materials and energy science, Graduate University of Science and Technology, Hanoi, Vietnam.*
[2]*Institute of Energy Science, Vietnam Academy of Science and Technology, IES – VAST, Hanoi, Vietnam.*
[3]*Infrastructure Development Center, FPT Telecom Joint Stock Company, Hanoi, Vietnam.*
[4]*Faculty of Electrical Engineering, The University of Danang – University of Science and Technology, Danang, Vietnam.*
**Corresponding author:* Ninh Nguyen Quang; Phone: +84-989-255-901; Email: nqninh@ies.vast.vn.

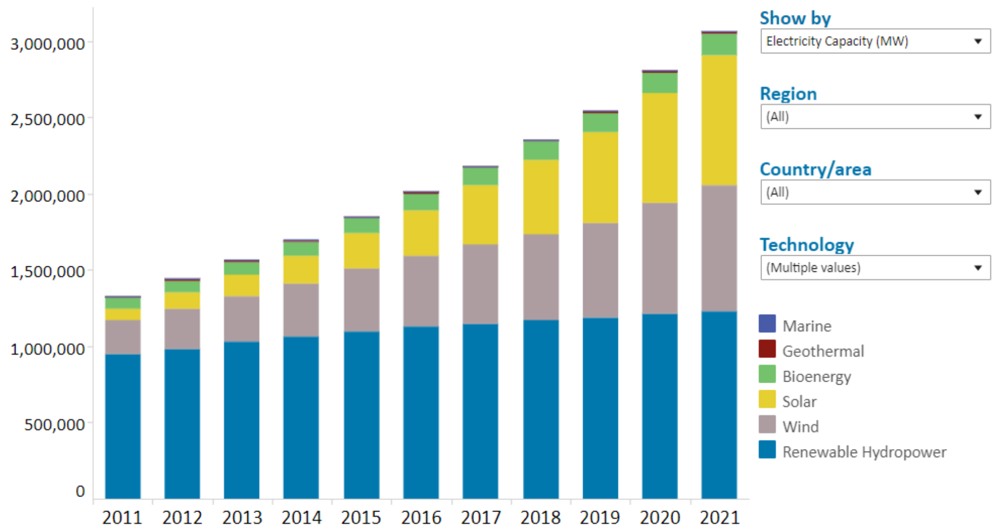


Fig.1. Global renewable generation capacity.

To evaluate the effectiveness of this predicting method, applying the obtained model to forecast the generating capacity of SPPs in other regions with different weather patterns is necessary. In this paper, the authors apply the training and data processing methods proposed in the article [21] to build a parameter matrix of the LSTM model for SPPs in the central part of Vietnam, where the climate is different from that of the southern part of Vietnam. The obtained model will be applied to predict the generating capacity of this SPP at specific climatic times of the region: sunny season, and rainy season. The input data are the real weather parameters and real generating capacity in the past of the SPP. Forecast results are compared with real output power data of the plant in the past.

Central Vietnam is an area characterized by two seasons: sunny and rainy. The rainy season usually starts from September to February next year. The application of the LSTM model for both sunny and rainy seasons is meaningful to survey the suitability of the application model. And the archived results could be the basis for proposing solutions to apply forecasting software for SPPs in Vietnam.

In this paper, the proposed LSTM method will be presented, then the input data will be processed to remove unnecessary outliers, the network training process, and the selection of the coefficient matrix of the LSTM in accordance with the requirements. The factory construction area is carried out. The experimental part will apply the model found to the scenarios: A typical day in the rainy season, A typical day in the sunny season, a typical month in the rainy season, and a typical month in the sunny season to evaluate the effectiveness of the obtained LSTM model. Finally, conclusions, evaluations, and future research directions are discussed.

2. MATERIALS AND METHODS

2.1. Long Short - term memory network

A neural network is a type of function that can map one set of values to another set in a mathematical way. It can be used in predictive models to convert feature vectors into scalar values, which is useful for solving regression problems. Recurrent neural networks (RNNs) are a type of neural network that can take input vectors as sequences, but they face a significant challenge with vanishing and exploding gradients, which is more severe than in traditional deep neural networks. This is because normal RNNs use the same weight parameters between recurrent units, unlike deep neural networks which have different weights between layers that can cancel each other out. To address this issue, the authors in [21] have investigated deep neural networks for long sequences, and one popular solution is a type of RNNs called Long Short-term Memory (LSTM). LSTM replaces every hidden unit of a normal RNN with LSTM cells, which can help mitigate the problem of vanishing and exploding gradients. Additionally, each LSTM cell has a special connection called the cell state, as shown in Figure 2.

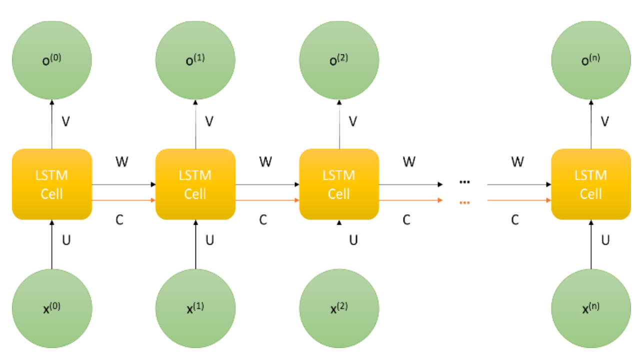


Fig.2. Long short-term memory (LSTM) networks.

2.2. Training with validation

To avoid overfitting, in [20], the authors used training techniques with validation. Part of the training dataset will be separated independently to continuously calculate the error after each epoch to make the decision about stopping or continuing training to achieve the model with the best accuracy as in Fig.3.

The training data in epochs (iterations) are loaded into the LSTM model with the default weighted matrix to find the training error value (calculated as the absolute value of the difference between the forecast generating power that the LSTM model and the weighted matrix of previous epoch with the input actual metering power).

After which, the weighted matrix is adjusted accordingly, consequently, the load verification dataset is also loaded into the LSTM model with the weighted matrix this epoch and the validation error value is found (calculated as the absolute error value of the result from the forecast generating power that the LSTM model and the new weighted matrix finds with the input data being the actual validation dataset).

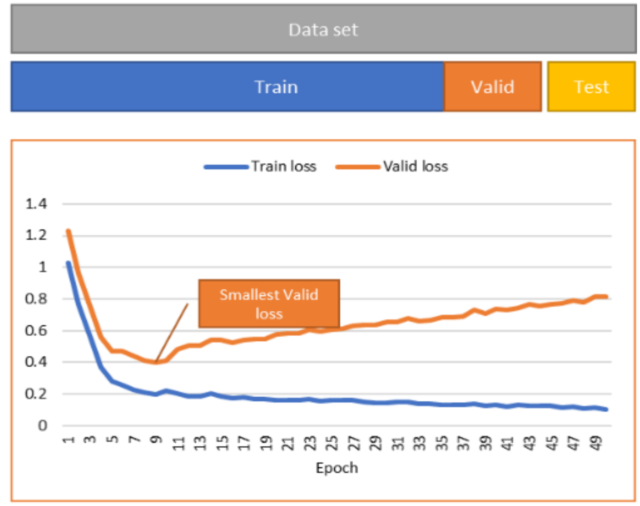
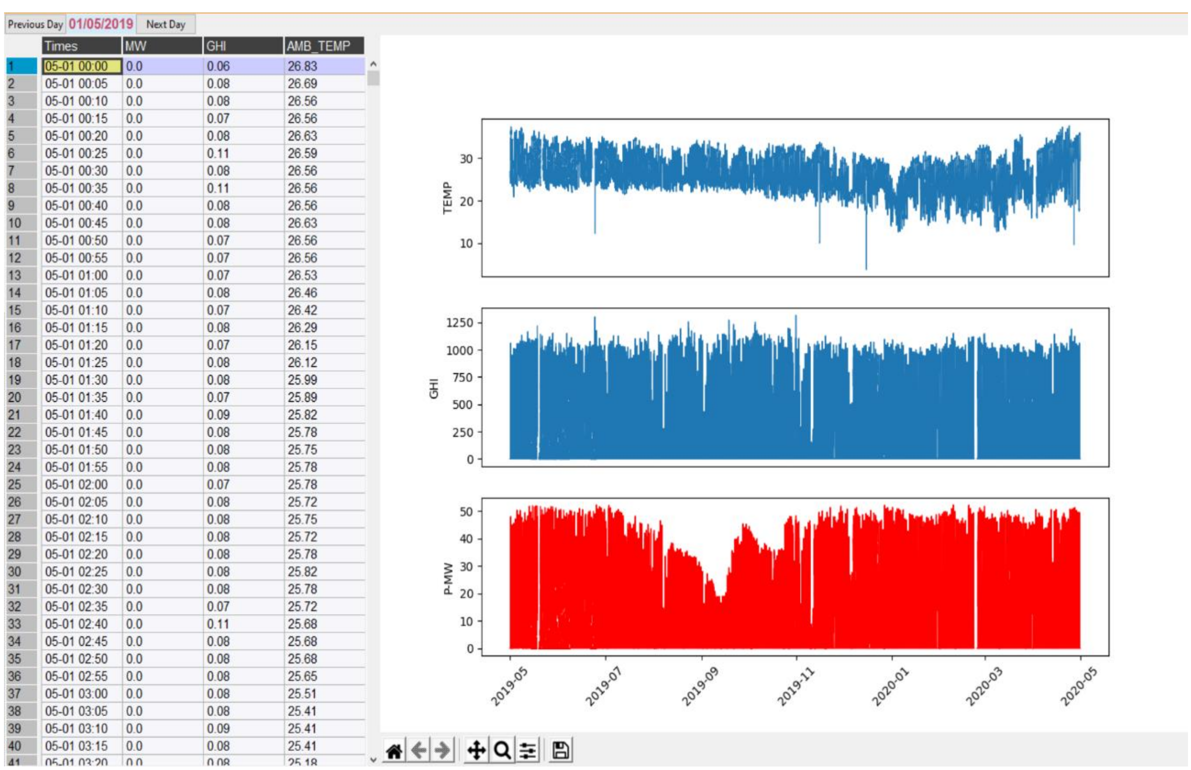


Fig.3. Train with validation: Dataset split & loss (MAE) during the training epochs.



	Times	MW	GHI	AMB_TEMP
1	05-01 00:00	0.0	0.06	26.83
2	05-01 00:05	0.0	0.08	26.69
3	05-01 00:10	0.0	0.08	26.56
4	05-01 00:15	0.0	0.07	26.56
5	05-01 00:20	0.0	0.08	26.63
6	05-01 00:25	0.0	0.11	26.59
7	05-01 00:30	0.0	0.08	26.56
8	05-01 00:35	0.0	0.11	26.56
9	05-01 00:40	0.0	0.08	26.56
10	05-01 00:45	0.0	0.08	26.63
11	05-01 00:50	0.0	0.07	26.56
12	05-01 00:55	0.0	0.07	26.56
13	05-01 01:00	0.0	0.07	26.53
14	05-01 01:05	0.0	0.08	26.46
15	05-01 01:10	0.0	0.07	26.42
16	05-01 01:15	0.0	0.08	26.29
17	05-01 01:20	0.0	0.07	26.15
18	05-01 01:25	0.0	0.08	26.12
19	05-01 01:30	0.0	0.08	25.99
20	05-01 01:35	0.0	0.07	25.89
21	05-01 01:40	0.0	0.09	25.82
22	05-01 01:45	0.0	0.08	25.78
23	05-01 01:50	0.0	0.08	25.75
24	05-01 01:55	0.0	0.08	25.78
25	05-01 02:00	0.0	0.07	25.78
26	05-01 02:05	0.0	0.08	25.72
27	05-01 02:10	0.0	0.08	25.75
28	05-01 02:15	0.0	0.08	25.72
29	05-01 02:20	0.0	0.08	25.78
30	05-01 02:25	0.0	0.08	25.82
31	05-01 02:30	0.0	0.08	25.78
32	05-01 02:35	0.0	0.07	25.72
33	05-01 02:40	0.0	0.11	25.68
34	05-01 02:45	0.0	0.08	25.68
35	05-01 02:50	0.0	0.08	25.68
36	05-01 02:55	0.0	0.08	25.65
37	05-01 03:00	0.0	0.08	25.51
38	05-01 03:05	0.0	0.08	25.41
39	05-01 03:10	0.0	0.09	25.41
40	05-01 03:15	0.0	0.08	25.41
41	05-01 03:20	0.0	0.08	25.18

2.3. Criteria for evaluating the forecasting results.

In this research, the criteria for evaluating the forecasting results are:

$$MAE = \frac{1}{N}\sum_{k=1}^{N}(P_M - P_P) \tag{1}$$

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{|P_M - P_P|}{P_{Rate}} \times 100\% \tag{2}$$

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(P_M - P_P)^2 \tag{3}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(P_M - P_P)^2} \quad (4)$$

where,

P_M - the measured capacity of the SPP
P_P - predicted capacity of the SPP
P_{Rate} - installed rated capacity of the SPP
N - the number of sampling points
MAE - the Mean Absolute Error
MAPE – the Mean Absolute Percentage Error
MSE – the Mean Square Error
RMSE - Root Mean Square Error

3. DATA

3.1. Collecting data

In this study, the operating data of a 50 MW solar plant in central Vietnam was collected. Dataset was collected from 5/2019 to 4/2020, Fig.4. The data resolution is 05 minutes. The features include:

- Average generating power in 05 minutes (MW)
- Average Global Horizontal Irradiance (GHI) radiation in 05 minutes (W/m^2)
- Average ambient temperature in 05 minutes (^{0}C)

One-month generating capacity of the plant during the sunny season, and the generating capacity on a typical day of the sunny season are presented in the Fig.5, Fig.6 respectively. The generating capacity for one month of the plant during the rainy season, and the generating power in a typical day of rainy season are presented in the Fig.7, and Fig.8 respectively.

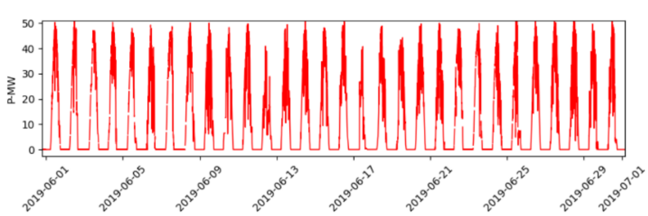


Fig.5. One-month generating capacity in the sunny season.

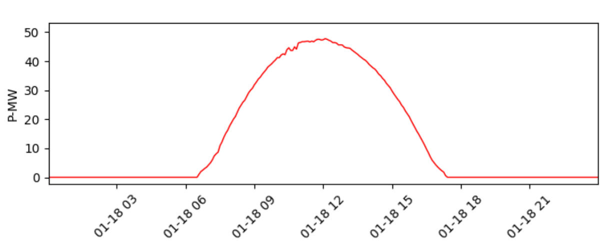


Fig.6. Generating capacity on a typical day of the sunny season.

The collected data has many unreliable points as shown in Fig.9. In many points even though the GHI is zero, the power P value is non-zero or the ratio between P and GHI is abnormal. Using the four-step method which uses P/GHI factor [20], to filter bad data, the relationship between generating power and irradiance after data processing is shown in Fig.10.

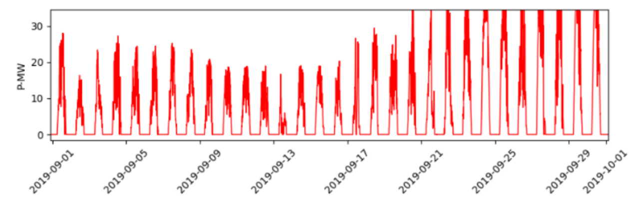


Fig.7. One-month generating capacity of the plant during the rainy season.

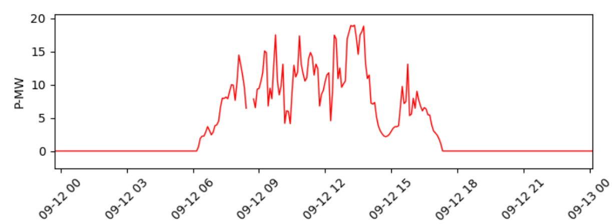


Fig.8. Generating capacity on a typical rainy season day

3.2. Data processing

The relationship between generating power and irradiance before data processing is shown in Fig.9.

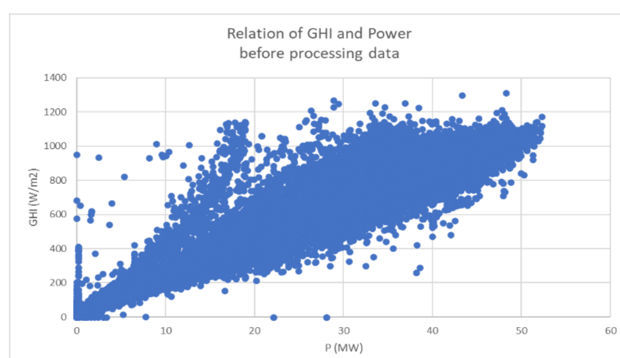


Fig.9. The relationship between generating power and irradiance before data processing.

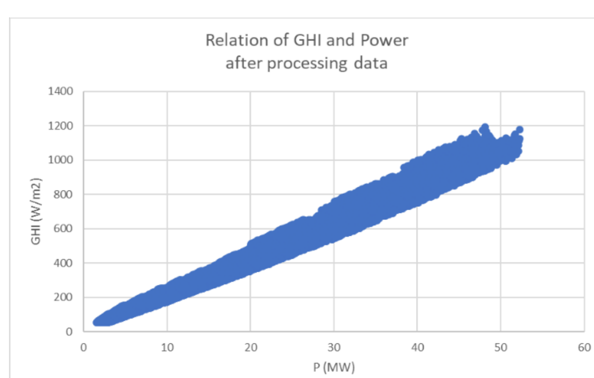


Fig. 10. The relationship between generating power and irradiance after data processing.

4. TRAINING THE MODEL

4.1. Model training parameter settings

The model has been trained by using the training techniques with validation being presented in section 2.2 with the parameters setting have proposed by [20] include:

- LSTM model 04 layers: each layer has 100 nodes
- Time delay input: 04 step
- Activation: ReLU
- Loss function: MAE (Mean Absolute Error)
- Optimization: Adam
- Valid rate: 10%
- Epoch train: 100 epochs,
- Early stopping: True, Patient = 20

4.2. Training process

Train loss and Valid loss graph of the training process is shown in Fig.11.

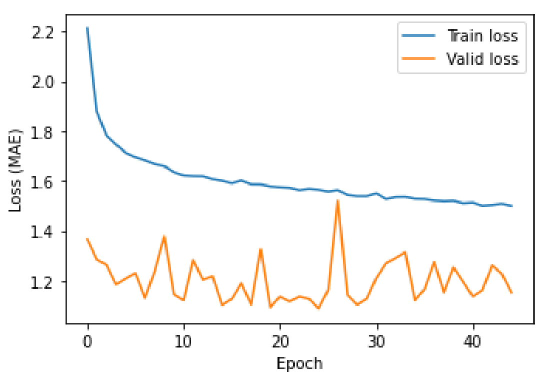


Fig.11. Train loss and Valid loss during training

The training uses the early stopping technique. From Fig.11 we can see that the valid loss min at epoch 25, after the next 20 epoch valid loss does not improve so the best model is selected at Epoch 25. The MAE of the best loss obtained is **1.09045.**

5. APPLYING AND RESULTS

The climate of Vietnamese central region belongs to the temperate type of climate. The characteristic of this type of climate is that there are 2 distinct seasons, the rainy season, and the sunny season. The rainy season starts in September, the heaviest rainfall intensity falls on October 10, November and ends the rainy season around January next year.

The authors chose October 15, 2020, as the typical day of the rainy season and June 10, 2020, as the typical day of the sunny season to forecast the generating capacity of the SPP.

5.1. Forecast for a day of the rainy season.

Applying the achieved model in section 4 for the selection date: October 15, 2020, the forecast results are shown in Fig.12.

In the rainy season, the generating capacity of the SPP is greatly affected by the cloud cover. The amount of radiation that the plant receives is quite low, leading to the shape of the generating output power is far from the bell-curve. The maximum capacity achieved at only a few times reaches above 40 MW, the rest mainly fluctuates around the threshold of 30 MW. The forecast result is moderately close to the actual shape and fluctuations of the day.

5.2. Forecast for a day of the sunny season.

Applying the achieved model in section 4 for the selection date: June 10, 2020, the forecast results are shown in Fig.13.

In the sunny season, the radiation conditions are good, the generating chart of the plant is close to the bell-curve shape. The generating capacity of the plant is relatively high with about 3 hours reaching over 40 MW. The results of the forecast model are comparatively close to the actual operating value.

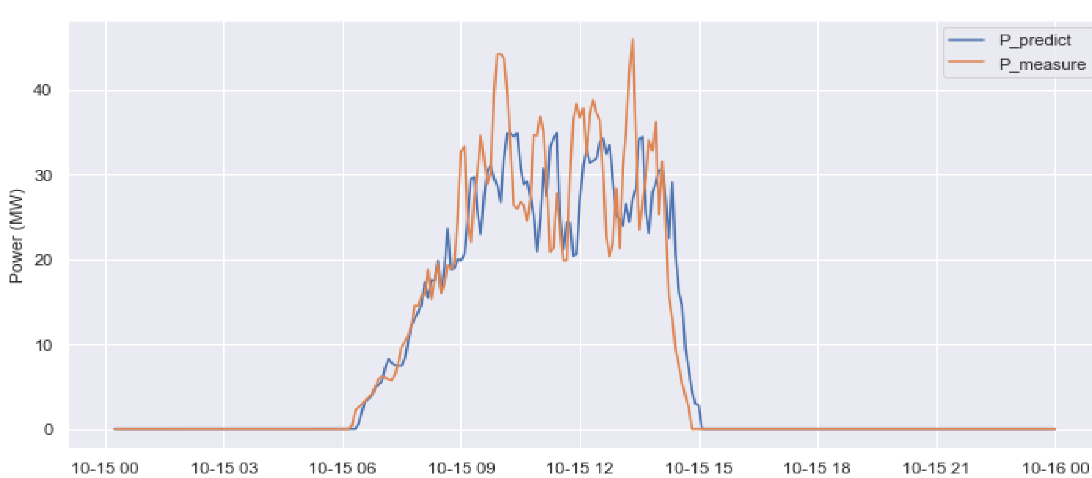


Fig.12. The forecast results of a day of rainy season.

Fig.13. The forecast results of a day of sunny season.

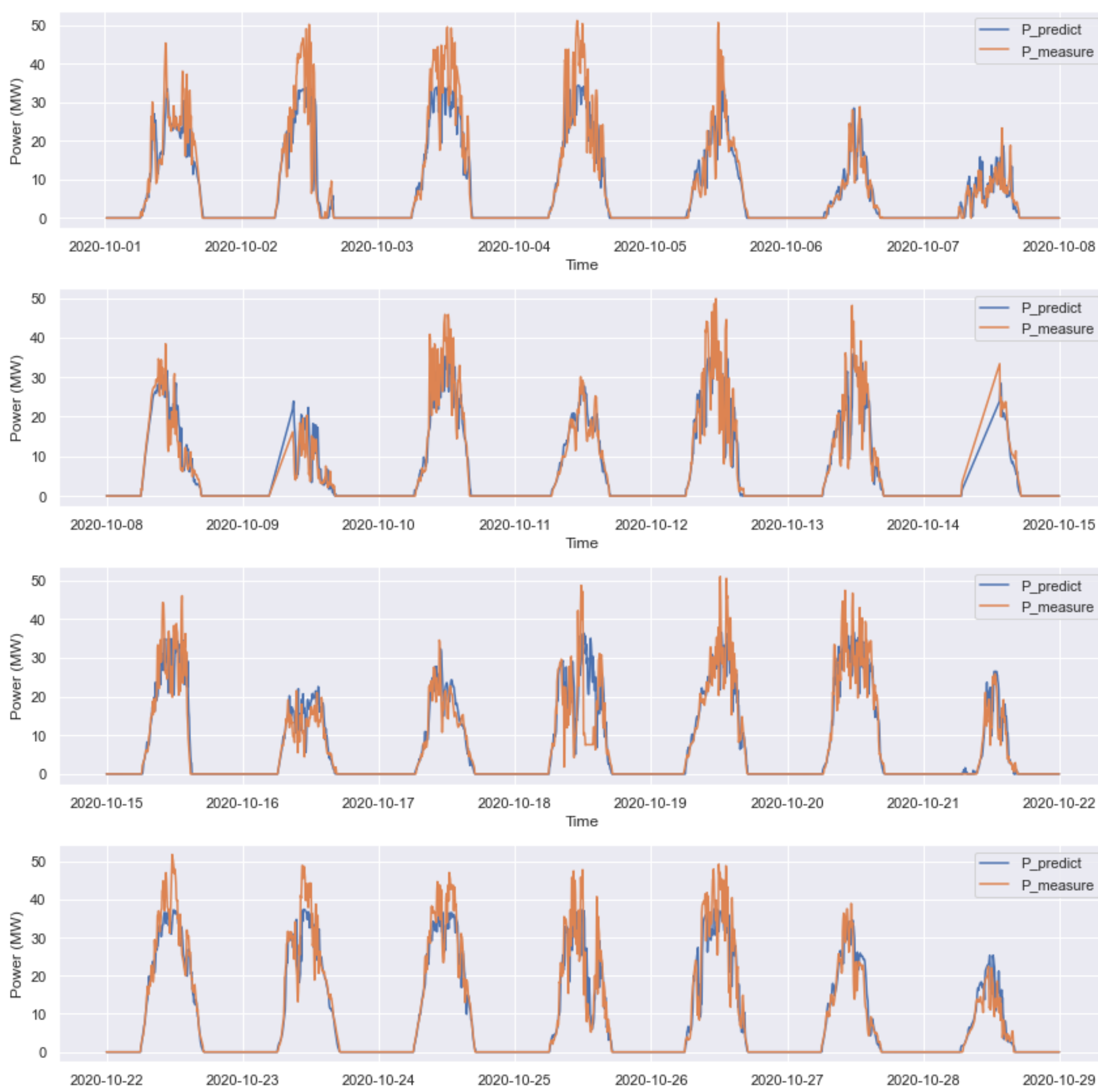


Fig.14. The forecast results of a month of rainy season.

5.3. Forecast for a month of the rainy season.

Applying the achieved model in section 4 for the selection month: October 2020, the forecast results are shown in Fig.14.

In October 2020, there were many days when the plant's generating capacity was at a very low level when the maximum generating capacity was only about 20 MW (equal to 40% of the rated capacity). The forecast model still gives relatively good results for these particularly low days.

5.4. Forecast for a month of the sunny season.

Applying the achieved model in section 4 for the selection month: June 2020, the forecast results are shown in Fig.15.

In June 2020, the plant's generating capacity was generally at a fairly high level, some days with a peak capacity of approximately 50 MW. With sunny days, the model performs very well, the error is quite small. For days with large drops, the forecast error results are acceptable.

Commonly metrics such as MAE, MAPE, MSE, and RMSE are calculated to evaluate the effectiveness of the forecast results. The summary of the results is shown in Table 1 below.

Through the result table, the model for forecast errors on actual historical datasets was relatively stable under experimental conditions:

- A rainy season day
- A sunny season day
- One month of the rainy season
- One month of the sunny season

Error of predictions:

- Full-day MAPE error (all-hour average) from about 4.0% - 4.7%
- MAPE error from 6:00-18:00 from about 8.0% - 9.4%
- RMSE error from 6:00-18:00 from about 6.1MW - 7.1MW

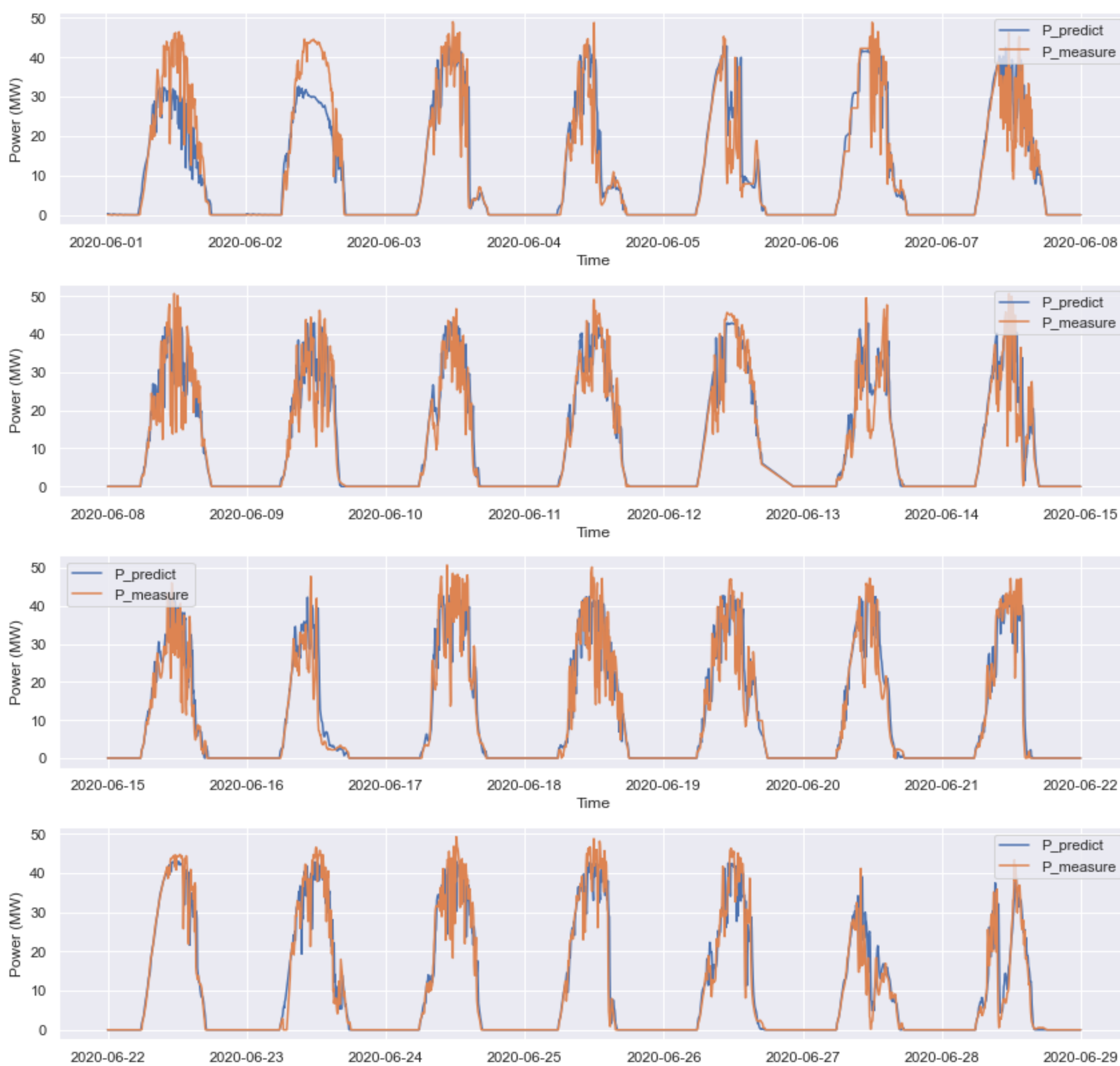


Fig.15. The forecast results of a month of sunny season.

- MAE error from 6:00-18:00 from about 4.0MW - 4.7MW
- MSE error from 6:00-18:00 from about 37.5MW^2 - 51.2MW^2

Table 1. Summary of forecast error evaluation results

	A day of rainy season	A day of sunny season	A month of rainy season	A month of sunny season
MAPE all day (%)	4.024	4.257	3.998	4.769
MAPE 6:00-18:00 (%)	7.993	8.436	8.006	9.402
RMSE 6:00-18:00 (MW)	6.272	6.123	6.314	7.152
MAE 6:00-18:00 (MW)	3.996	4.218	4.00	4.701

6. CONCLUSIONS

Solar Power Plant Output Forecasting Model Based on the LSTM Method Using Validation in Different Seasons of the Year in Vietnam still shows effectiveness. With another SPP, of the same scale, different geographical location, the authors applied the training method proposed by [20] to obtain the weighted matrix of LSTM model, and in different weather areas, different seasons: rainy season and sunny season, the forecast results achieved were relatively good with forecast errors of less than 10%. In the next research direction, the forecast model will be further tested in different SPPs, finally, the model of the system forecasting the generating capacity for a SPP will be researched and implemented in practice.

ACKNOWLEDGEMENTS

The authors especially thank the Vietnam Academy of Science and Technology (VAST), Institute of Energy Science (IES), and Graduate University of Science and Technology (GUST) for providing funding and generating interest in the Project "Research on methodology and develop software to short-term forecast the output of solar power plants based on Artificial Intelligence", VAST07.01/21-22.

REFERENCES

[1] International Renewable Energy Agency. 2021. https://www.irena.org/solar. https://www.irena.org/solar

[2] Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. 2020. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. In Renewable and Sustainable Energy Reviews, Vol. 124. Elsevier Ltd. https://doi.org/10.1016/j.rser.2020.109792

[3] Yin, S., Wang, J., Li, Z., & Fang, X. 2021. State-of-the-art short-term electricity market operation with solar generation: A review. In Renewable and Sustainable Energy Reviews, Vol. 138. Elsevier Ltd. https://doi.org/10.1016/j.rser.2020.110647

[4] D. D. Le. 2019. An Enhancement to Time Series Model for Short-Term Forecasting of Multi-Regional Electricity Load. GMSARN International Journal 13, 202 - 207.

[5] N. Phuangpornpitak and W. Prommee. 2016. A Study of Load Demand Forecasting Models in Electric Power System Operation and Planning. GMSARN International Journal 10, 19 - 24.

[6] Thi Hoai Thu Nguyen, Quoc Bao Phan, Vu Nhat Nam Nguyen, Hai Minh Pham. 2021. Day-ahead electricity load forecasting based on hybrid model of EEMD and Bidirectional LSTM. Proceedings of ICFNDS 2021: The 5th International Conference on Future Networks & Distributed Systems. December 2021. Pages 31–41https://doi.org/10.1145/3508072.3508079

[7] Dutta, S., Li, Y., Venkataraman, A., Costa, L. M., Jiang, T., Plana, R., Tordjman, P., Choo, F. H., Foo, C. F., & Puttgen, H. B. 2017. Load and Renewable Energy Forecasting for a Microgrid using Persistence Technique. Energy Procedia, 143, 617–622. https://doi.org/10.1016/j.egypro.2017.12.736

[8] Zameer, A., Arshad, J., Khan, A., & Raja, M. A. Z. 2017. Intelligent and robust prediction of short-term wind power using genetic programming-based ensemble of neural networks. Energy Conversion and Management, 134, 361–372. https://doi.org/10.1016/j.enconman.2016.12.032

[9] Zhao, J., Guo, Z. H., Su, Z. Y., Zhao, Z. Y., Xiao, X., & Liu, F. 2016. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. Applied Energy, 162, 808–826. https://doi.org/10.1016/j.apenergy.2015.10.145

[10] Montgomery, D.C.; Jennings, C.L.; Kulahci, M. 2008. Introduction to Time Series Analysis and Forecasting, 1st ed.; Wiley: Hoboken, NJ, USA.

[11] Reikard, G. 2009. Predicting solar radiation at high resolutions: A comparison of time series forecasts. Sol. Energy, 83, 342–349. https://doi.org/10.1016/j.solener.2008.08.007

[12] C. Paolik, C. Voyant, M. Muselli, and M. Nivet. 2009. Solar radiation forecasting using ad-hoc time series preprocessing and neural networks. Proceedings of. 5th International Conference on Emerging Intelligent Computing technology and applications, Ulsan, South Korea, pp. 898-907.

[13] Mellit, A.; Massi Pavan, A. 2010. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Sol. Energy, 84, 807–821. DOI: 10.1016/j.solener.2010.02.006

[14] Mellit, A.; Shaari, S. 2009. Recurrent neural network-based forecasting of the daily electricity generation of a photovoltaic power system. Ecolog. Veh. Renew. Energy 2009, 2–7.

[15] Mellit, A.; Shaari, S. 2009. Recurrent Neural Network-Based Forecasting of The Daily Electricity Generation of a Photovoltaic Power System. In Proceedings of the Ecological Vehicle and Renewable Energy, Monte-Carlo, Monaco, 26–29 March 2009; pp. 26–29.

[16] Mellit, A.; Saglam, S.; Kalogirou, S.A. 2013. Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renew. Energy 2013, 60,

71–78. https:// doi.org/10.1016/j.renene.2013.04.011

[17] Raza, M. Q., Nadarajah, M., & Ekanayake, C. 2016. On recent advances in PV output power forecast. Solar Energy, 136, 125–144. DOI:10.1016/j.solener.2016.06.073

[18] Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. 2018. Solar photovoltaic generation forecasting methods: A review. Energy Convers. Manag. 2018, 156, 459–497. https://doi.org/ 10.1016/j.enconman.2017.11.019

[19] Nguyen, N. Q., Duy Bui, L., Doan, B. Van, Sanseverino, E. R., Cara, D. Di, & Nguyen, Q. D. 2021. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. Electric Power Systems Research, 199, 107427. https://doi.org/10.1016/j.epsr.2021.107427

[20] Linh, Duy Bui, Ninh Quang Nguyen, Binh Van Doan & Sanseverino, E. R. 2022. Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process, a case study in Vietnam. Electric Power Systems Research, 213, 108706. https://doi.org/10.1016/J.EPSR.2022.108706

[21] Hochreiter, S., and Schmidhuber, J. 1997. Long Short-Term Memory. Neural Computation, 9(8). https://doi.org/10.1162/ neco.1997.9.8.1735.