Research Article

A Short-Term Photovoltaic Power Generation Forecast Method Based on LSTM

Yang Li,[1] **Feng Ye** [iD]**,**[1] **Zihao Liu,**[2] **Zhijian Wang,**[1] **and Yupeng Mao**[3]

[1]*School of Computer and Information, Hohai University, Nanjing, China*
[2]*China Telecom Nanjing Branch, Nanjing, China*
[3]*Nanjing Yiting Internet of Things Technology Co., Ltd., Nanjing, China*

Correspondence should be addressed to Feng Ye; yefeng1022@hhu.edu.cn

The intermittence and fluctuation of photovoltaic power generation seriously affect output power reliability, efficiency, fault detection of photovoltaic power grid, etc. The precise forecasting of photovoltaic power generation is the critical method to solve the above limitations. Current photovoltaic power generation forecasting methods generally usually adopt meteorological data and historical continuous photovoltaic power generation as inputs, but they do not take into account historical periodic photovoltaic power generation as inputs, which makes the existing methods inadequate in learning time correlation. Therefore, to further study the time correlation for improving the prediction accuracy, an LSTM-FC deep learning algorithm composed of long-term short-term memory (LSTM) and fully connected (FC) layers is proposed. The double-branch input of the model enables it not only to consider the impact of meteorological data on power generation but also to consider time continuity and periodic dependence, thereby improving the prediction accuracy to a certain extent. In this paper, meteorological data, historical continuous data, and historical periodic data are used as experimental data, and these three types of data are combined into different input forms to evaluate and compare LSTM-FC with other baseline models, including support vector machines (SVM), gradient boosting decision tree (GBDT), generalized regression neural network (GRNN), feedforward neural network (FFNN), and LSTM. The simulation results show that the accuracy of the models with meteorological data, continuous data, and periodic data as input is higher than that of other input forms, and the accuracy of LSTM-FC is the highest among these models, and its root mean square error (RMSE) is 11.79% higher than that of SVM.

1. Introduction

The gradual depletion and pollution of resource-based energy have become the focus of attention of the whole society, while renewable energy represented by solar energy has absolute security, cleanliness, and resource sufficiency and plays an important role in the long-term energy strategy [1, 2]. According to the statistical data results of the International Renewable Energy Agency (IRENA) in 2019, China's installed photovoltaic (PV) power generation capacity has reached 205.493 GW [3]. PV power stations are based on solar energy. However, as the output of PV power generation is highly fluctuant and intermittent, when PV power generation reaches a certain threshold, it will bring huge impact and challenges to the power system and reduce

the efficiency and reliability. The purpose of this paper is to propose a deep learning algorithm to accurately predict the PV power generation, so as to reduce the impact of the fluctuation and intermittence of PV power generation on the PV system, improve the operation efficiency, generate stable and reliable power supply, and help the operators adjust the operation mode and decision implementation.

At present, many modern machine learning models including some deep learning models have been well applied in the fields related to PV power generation prediction [4]. In addition, several hybrid model based methods are also proposed for improving prediction accuracy [5–7]. It can be concluded from [8–11] that traditional machine learning methods often require a lot of feature engineering work before training and testing, which makes the preliminary

work relatively heavy. On the other hand, these methods consider fewer relevant factors, such as only considering meteorological data. In [12–18], some deep learning models, including LSTM, GRNN, and FFNN, have also been applied in the field of PV power generation. However, they only consider some meteorology data and historical continuous data, regardless of historical periodic data, so there is still room for improvement in accuracy.

To compensate for these shortcomings, our contribution is that we use real public historical PV power generation data and meteorological information to forecast PV power generation through LSTM-FC in the next 1 hour. LSTM models have performed well in the field of temporal modelling, while the fully connected layers usually are suitable for feature mapping. An LSTM-FC structure with two branches, including main input and auxiliary input, not only use LSTMs layer to obtain time continuity and periodicity, but also obtain the important role of meteorological information through FCs layer. This indicates that these two branches play complementary roles, which can further improve the accuracy of PV power generation prediction.

The rest of this paper is organized as follows. The second section mainly introduces the related work to this paper. The third section mainly introduces the model structure and evaluation metrics. In the fourth section, continuous data, periodic data, and meteorological data are used as experimental data to obtain the prediction results of LSTM-FC and baseline model, and their errors were compared using RMSE, etc. Finally, the summary and future prospects are given.

2. Related Work

The current machine learning methods including some deep learning methods can be widely used in the short-term forecasting of PV power generation. For example, Shi et al. developed a method based on SVM to predict PV power generation under different weather conditions [10]. They divided four models according to different weather conditions and then used historical data and weather forecasts as inputs to predict PV power generation. The authors obtained promising results when testing their technique to forecast the power output of a PV station in China. However, the data sets in their paper are mostly sunny or rainy, so it is troublesome to design different models for regions with more weather types.

Wang et al. proposed a short-term forecasting method of PV power based on GBDT [11]. They used historical weather data and PV power data to train the model. The simulation results show that this method is superior to SVM and autoregressive moving average model (ARMA). However, this method does not use periodic PV power data and has few features. Ramsami and Oree proposed a prediction method based on hybrid model [15]. First, stepwise regression (SR) was used to obtain important features as input, and then FFNN, GRNN, and multiple linear regression (MLR) were used to predict. Simulation results showed that this method could achieve the same accuracy or even better with a small number of features. For the same problem, this

method only considers the average value of meteorological data of the day before yesterday and does not carry out SR processing for relevant characteristics in time. In [16], Cheng et al. proposed PV power forecast research of PV system based on double-level neural network; the input of the double-level neural network adopts the calculated value or the predicted value and learns the existing measured value through the network to correct the theoretical value and get the prediction result. This idea can use LSTM to further improve the prediction accuracy.

Huang et al. proposed a short-term prediction-based solar irradiation on LSTM-MLP [17]. In this method, the historical continuous irradiance and the meteorological data of the predicted day are taken as the input of LSTM, and then the output of LSTM and meteorological data are combined as the input of MLP, and finally the final result is obtained through MLP. The simulation results show that the performance of the model is better than that of the baseline model. In [18, 19], the authors also use prediction methods based on LSTM or other hybrid models of deep learning models and LSTM, and the simulation results also prove that these methods improve the accuracy of prediction on the original basis. However, they only considered the correlation between continuity data and did not take the correlation of data in the period as input, so there was room for the prediction accuracy to rise.

In a word, the aforementioned methods only consider the impact of the historical continuous PV power generation and meteorological data on the power generation, and they do not consider the time-periodic dependence. In LSTM-FC model, it can obtain not only the effect of meteorological data to PV power generation, but also time continuity and periodicity dependence. Next, we'll look at the model structure of LSTM-FC and why it has the above characteristics.

3. The Proposed Model

3.1. Model Structure of LSTM-FC. LSTM can well solve the problem of gradient disappearance and gradient explosion and has a good application for the acquisition of time correlation, while FC can better acquire the mapping relationship between features. We can take advantage of the complementary characteristics of the two networks to further improve the accuracy of the prediction. Below, we will introduce the LSTMF-FC network structure.

As shown in Figure 1, this is the model structure of LSTM-FC proposed in this paper. It contains two components. The first component consists of an input layer, two LSTM layers, and an output layer; it has two input types A and B. For input A, the historical PV power generation at $[t-1, t-2, \ldots, t-n]$ and $[t-1*p, t-2*p, \ldots, t-n*p]$ time is the feed LSTM layer1, and then go through two LSTM nodes and the output layer to get the output of the first component, where n is the length of the time series (number of samples) and p is the period interval. The input of B is one more meteorological data than that of A. The second component consists of an input layer, three fully connected layers, and an output layer. The input layer is the
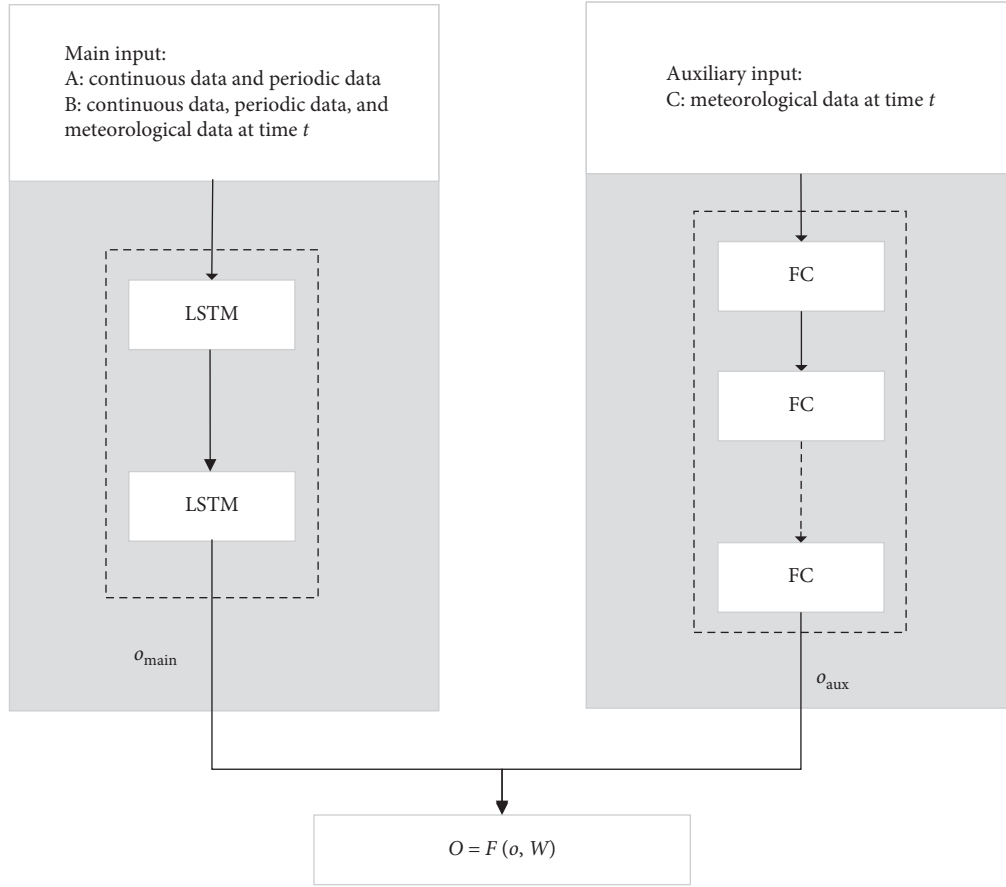
Figure 1: The structure of LSTM-FC.

meteorological data at time t, and then the data is passed to the FC layer1, and then go through several hidden layers and the output layer to get the output of the second component. The final prediction result of the two components will be weighted and summed according to the output of the two components in a certain proportion.

The expression of the operation in Figure 1 is as follows:

$$
\begin{aligned}
h_{\text{LSTM}^{(1)}} &= y_{\text{LSTM}^{(1)}}\left(x_{\text{main\_input}}\right), \\
o_{\text{main}} &= y_{\text{LSTM}^{(n)}}\left(h_{\text{LSTM}^{(n-1)}}\right), \\
h_{fc^{(1)}} &= y_{fc^{(1)}}\left(x_{\text{aux\_input}}\right), \\
o_{\text{aux}} &= y_{fc^{(n)}}\left(h_{fc^{(n-1)}}\right), \\
O &= F(o, W) = W_{\text{main}} * o_{\text{main}} + W_{\text{aux}} * o_{\text{aux}},
\end{aligned}
\tag{1}
$$

$x_{\text{main\_input}}$ represents the main input, such as input A or input B. y_{LSTM} represents the LSTM layer, h_{LSTM} represents the output passing through the LSTM layer, and $x_{\text{aux\_input}}$ represents the auxiliary input, which is the meteorological data at time t. o_{main} represents the output of the first component, and o_{aux} represents the output of the second component. The final output result is fused by the parameter matrix, and W is learned by the model itself.

As shown in Figure 1, the models can be divided into Model-A, Model-B, Model-AC, and Model-BC according to different inputs. In the following sections, we will discuss different inputs.

how to choose an appropriate length of contiguity data sequence and the sequence length of periodicity data and then get an excellent model from these models.

3.2. Evaluation Function of Model Accuracy.
To better distinguish the performance of the model and implement a baseline to compare with other baseline models, some evaluation indicators need to be set. The evaluation functions used to verify the model performance are presented below, including RMSE, standardized root mean square error (nRMSE), mean absolute error (MAE), and determination coefficient (R^2).

$$
\begin{aligned}
\text{RMSE} &= \sqrt{\frac{1}{n}\sum_{i}^{n}\left(Y_i - Y_i'\right)^2}, \\
n\text{RMSE} &= \frac{\text{RMSE}}{\mu_Y}, \\
R^2 &= 1 - \frac{(1/n)\sum_{i=1}^{n}\left(Y_i - Y_i'\right)^2}{(1/n)\sum_{i=1}^{n}\left(Y_i - \mu_{Y'}\right)^2} = 1 - \frac{\sum_{i=1}^{n}\left(Y_i - Y_i'\right)^2}{\sum_{i=1}^{n}\left(Y_i - \mu_{Y'}\right)^2},
\end{aligned}
\tag{2}
$$

Y is the predicted value obtained by the model, and Y' is the expected true value. $\mu_{Y'}$ is the mean of the expected values.

Each evaluation index has its own specific target. For PV power generation, RMSE, nRMSE, and MAE can well reflect the dispersion degree between the predicted value and the real value, but in some cases, R^2 is more useful than either of the above and can better explain the performance of the model.

4. Experiment and Result Analysis

4.1. Experimental Data and Environment. The experimental data is real application data, which comes from Jinan PV operation and maintenance platform. Considering the confidentiality of data, the PV power generation data and meteorological data is from 2018-1-1 to 2018-12-31. The meteorological data includes temperature, humidity, weather, wind direction, and wind speed. Since the PV system does not generate electricity during certain periods, the time set in this article is 7–19 hours a day. As shown in Table 1, this is the structure of the data. To meet the requirements of training, we should preprocess the data, such as data cleaning and data normalization. The specific steps are as follows. Firstly, weather and wind direction need to be encoded into data that computers can process. After processing, the corresponding number for local cloudy days is 15 and the corresponding number for east wind is 3.

Secondly, the outliers need to be processed to reduce the impact on the accuracy of the prediction algorithm. Currently, the main outlier detection algorithms include One-ClassSVM, EllipticEnvelope, IsolationForest, and LOF [20]. Here, the four methods above are, respectively, used to test the data, and it is found that LOF is the most suitable method compared with the other three methods. As shown in Figure 2, this is the outlier obtained by the LOF method; x-axis is the sample index, and y-axis is the hourly PV power generating. For the removal of outliers, the method used is to make up with the PV power generated at the same time, the day before or the day after the outlier at that time. The processed data is shown in Figure 3, which shows that the processed data is normal.

Thirdly, we analyze the characteristics and calculate the correlation between them and PV power generation. The Pearson correlation coefficient is used to measure whether two data sets are on a line, and it is used to measure the linear relationship between distance variables. Its formula is shown in formula (3). To determine whether meteorological data should be taken as input features, the meteorological data and PV power generation data are calculated in this paper. Table 2 shows the Pearson correlation coefficient between these five meteorological variables and PV power generation. x is the meteorological data, y is PV power generation, and u is the average value.

$$p_{x,y} = \frac{\sum (x - \mu_x) * (y - \mu_y)}{\sqrt{\sum (x - \mu_x)^2 * \sum (y - \mu_y)^2}}. \quad (3)$$

It can be observed that there is almost no correlation between the weather data and the PV power generation, and there is a certain correlation between the other several variables, among which temperature and humidity are the highest. Therefore, in addition to the weather variable, several other variables can be used as input variables.

Figure 4 shows the continuity of PV power, and the change in continuous time is relatively smooth. As shown in Figure 5, it is obvious that PV power reaches its peak at noon every day, and it is lower in the morning and evening. The curve shows an obvious periodicity. Next, we analyze the Pearson correlation coefficient between historical continuous PV power generation and periodic PV power generation and the PV power generation on the forecast day. Figure 6 shows the fitting degree of the PV power at time $t - 1$ and $t - 2$ and t, indicating that there is a great dependence between the two. Figure 7 shows the fitting degree of the PV power at time $t - 1 * p$ and $t - 2 * p$ and t, and it can also be seen that there is a great dependence between the two. We only analyzed the correlation between the data with a lagging time of 2 hours (or period), but longer lagging time is still relevant. Therefore, we will conduct experiments to determine the optimal lagging time later.

After the above process is completed, we finally form a data set which includes 13,792 training data and 3,412 test data. As shown in Table 3, all the data are integrated into a table, and the data will be constructed according to different input requirements when reading the data. Since the characteristic dimensions cannot be fully displayed here, we only show meteorological data and a small amount of continuous data and periodic data in the table, in which power is PV power generation at time t, c_i is PV power generation at time $t - i$, and p_i is PV power generation at time $t - i * p$. The statistical results of some characteristics of training data are shown in Table 4, including mean value, standard deviation, minimum value, and maximum value.

In this experiment, the hardware environment includes Windows 10, 6G memory, and GTX 1660 Ti graphics card. The software environment includes CUDA 9.0, Cudnn7.4.1, scikit-learn0.22.2, Python3.6, and Keras2.1.6, and the corresponding Keras back end is TensorFlow 1.11.0.

4.2. Forecasting Test. In Figure 1, we divided four different models according to the different input types and the choice of double branches, namely, model-A, model-B, model-AC, and model-BC, but each model contains continuous and periodic PV power. Therefore, the length of the continuous data series and the length of the periodic data series should be determined first. According to these two kinds of data, the input variables of model-A are formed. In addition to these inputs, meteorological data should be added to model-B, model-AC, and model-BC. Then, these models are compared and evaluated to verify the influence of different input types on the accuracy of model prediction. Finally, the best model is selected according to the optimal evaluation results.

The following is to conduct continuous training to determine the optimal length of continuous PV power and periodic PV power, respectively. Considering that the effective generating time of one day obtained in the previous data is 7–19 hours, the maximum length of continuous data

TABLE 1: The structure and form of data.

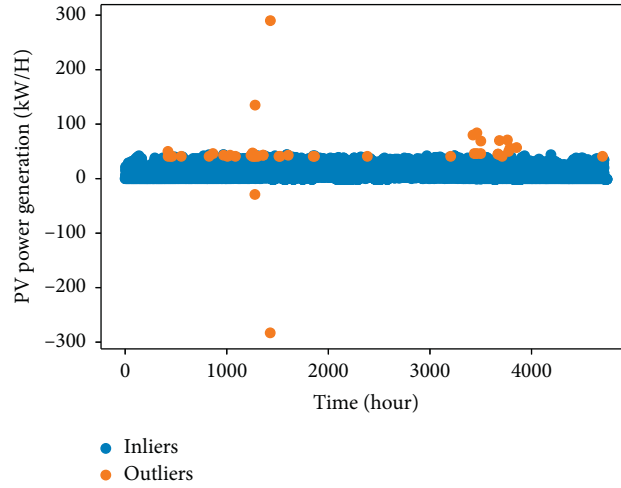| Time	Power	Weather	Humidity	Temperature	Wind speed	Wind direction
2018-01-01 07	0	Local cloudy	32	−1	4	The east wind
2018-01-01 08	0	Local cloudy	32	−1	4	The east wind
2018-01-01 09	2.0	Local cloudy	28	1	4	The east wind
2018-01-01 10	7.0	Local cloudy	23	3	5	The east wind
2018-01-01 11	15.0	Local cloudy	18	5	5	The east wind

● Inliers
● Outliers

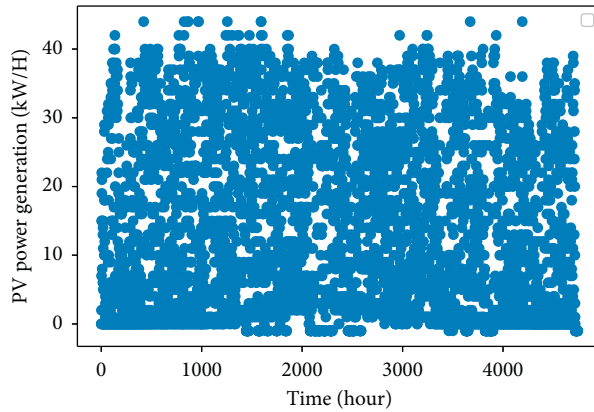FIGURE 2: Outliers of data (not cleaned up).

FIGURE 3: Outliers of data (cleaned up).

TABLE 2: Pearson's correlation coefficient between meteorological data and PV power generation.

Meteorological parameters	Pearson's correlation coefficient
Temperature	0.2589
Humidity	−0.3686
Wind speed	0.1915
Weather	0.0114
Wind direction	0.1657

is set to the first 9 hours. Considering the time consumption and the decrease of long time series dependence, the maximum sequence length of periodic data is also 9.

After many training sessions, as shown in Figure 8(a), RMSE decreases as the length of the input variable increases at an interval of 1 hour. It can be seen that the
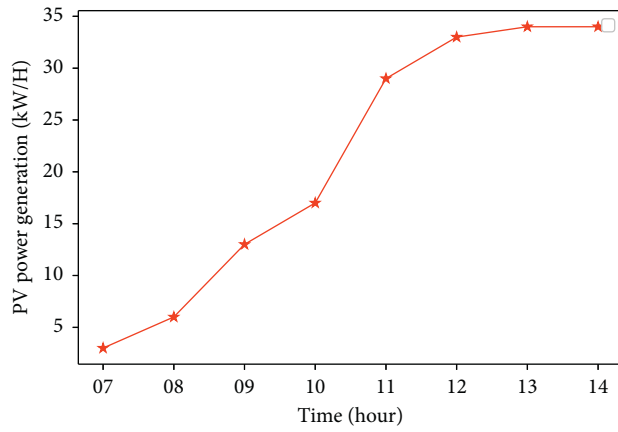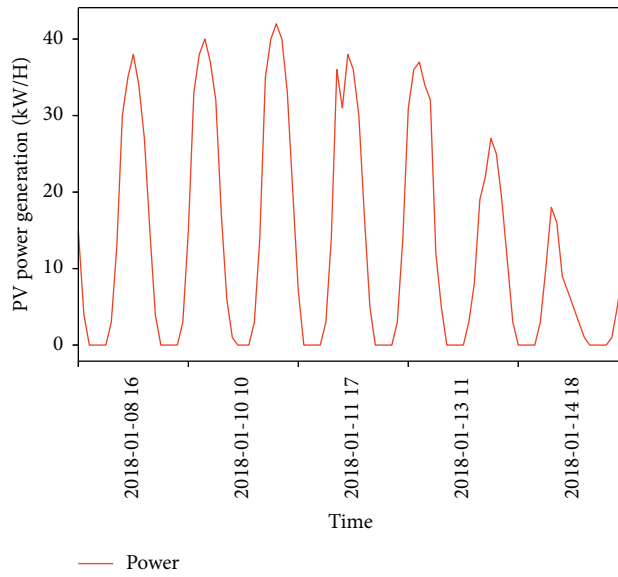
FIGURE 4: Contiguity of PV power generation.
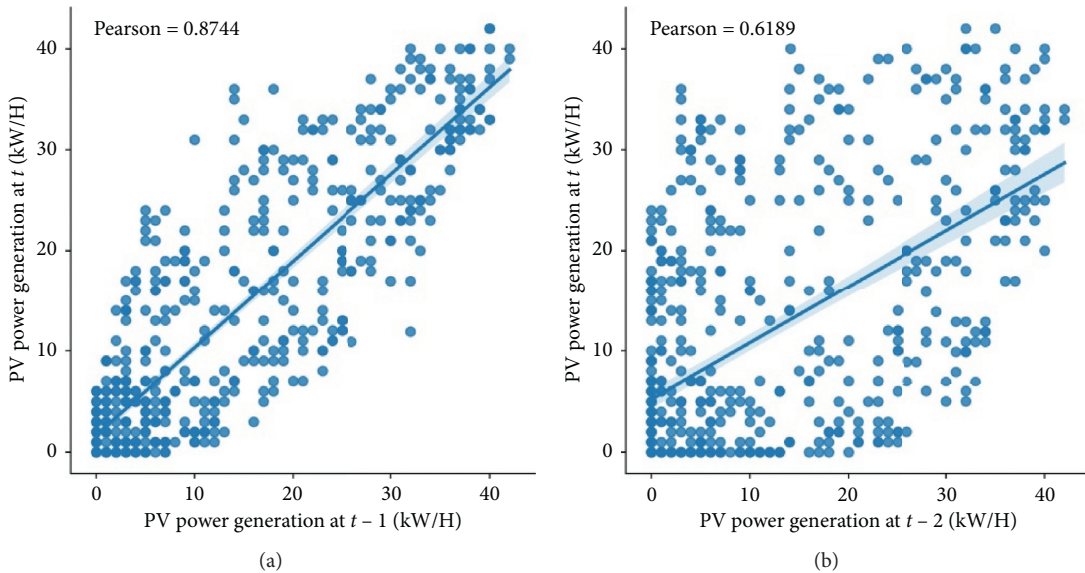

FIGURE 5: Periodicity of PV power generation.

(a)

(b)

FIGURE 6: Scatter plot of PV power generation at time $t - 1$ and $t - 2$ vs. t.
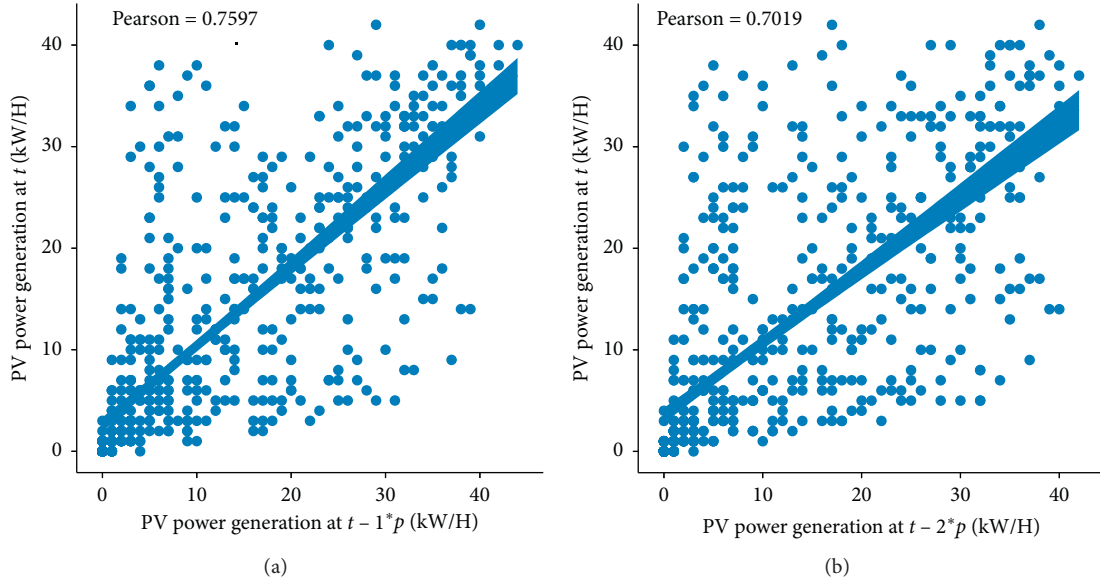
(a)

(b)

FIGURE 7: Scatter plot of PV power generation at time $t - 1 * p$ and $t - 2 * p$ vs. t.

TABLE 3: The sample of training and testing data.

Temperature	Humidity	Wind speed	Wind direction	c_1	c_2	c_3	p_1	p_2	p_3	Power
−3	15	11	6	15	3	0	30	24	3	15
−3	14	11	6	33	15	3	35	33	3	22
−2	14	10	6	38	33	15	38	35	5	19
−2	13	10	6	40	38	33	34	35	6	21

TABLE 4: The statistical results of training data.

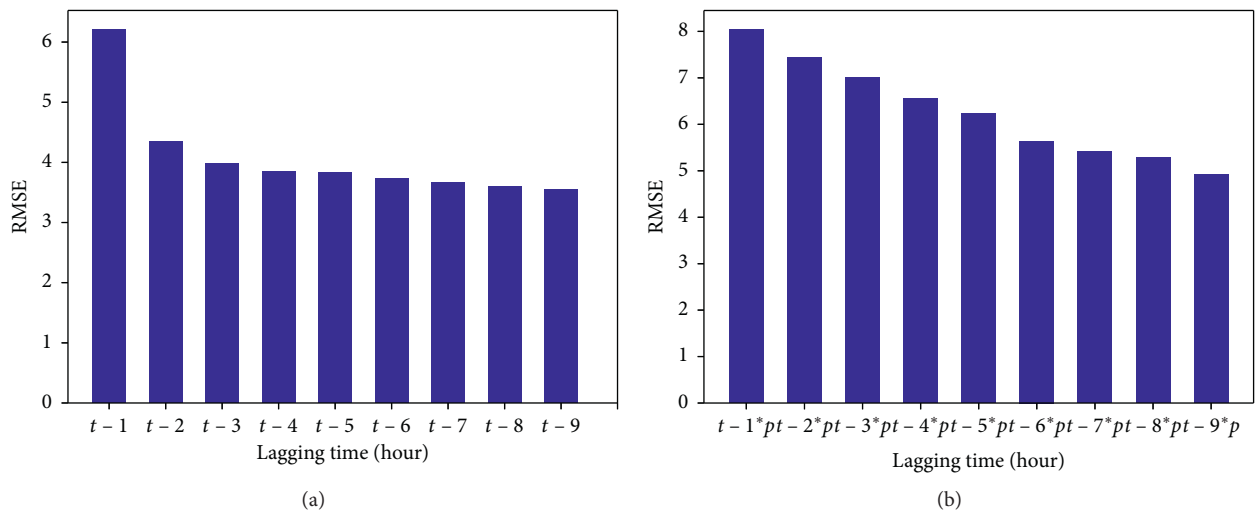|	Power	Humidity	Temperature
Mean	14.47	39.41	15.00
Std	13.03	19.52	13.16
Min	0.00	7.00	−12.00
Max	46.00	97.00	40.00

(a)

(b)

FIGURE 8: Error of different lengths of continuous (periodic) data sequence on prediction. (a) RMSE of different length of contiguity data sequence. (b) RMSE of different length of periodicity data sequence.

TABLE 5: The influence of different inputs on the prediction accuracy of the model.

	nRMSE	RMSE	R^2
Model-A	0.2209	3.2453	0.9357
Model-B	0.1919	2.8189	0.9515
Model-AC	0.2009	2.9515	0.9468
Model-BC	0.1743	2.5605	0.9600

TABLE 6: LSTM-FC parameters setting.

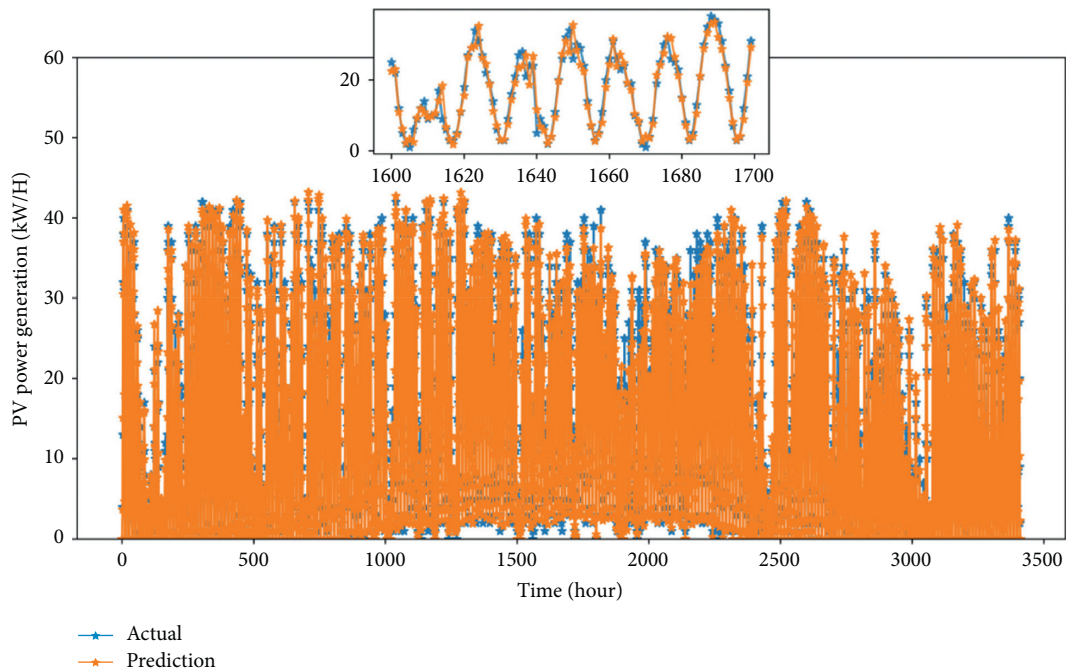| Parameters	Content	Setting
n_LSTM	The number of LSTMs	2
n_FC	Th number of FCs	4
batch_size	The number of batch training samples	64
P	Periodic time interval	13
Adam	Optimization algorithm	—
Relu	Activation function	—
Lr	Learning rate	0.001
Lc	The number of contiguity samples	4
Lp	The number of periodic samples	9
epoch	Iterations	200

FIGURE 9: Prediction results of LSTM-FC on test set.

prediction accuracy is the highest when the lagging time is 9, but the decline is no longer significant after the lagging time is 4. As shown in Figure 8(b), RMSE decreases as the length of the input variable increases, and the prediction accuracy is highest when the lagging time is 9, where p is 13. After the above analysis, the length of the final selected continuous data is 4, and the length of the periodic data is 9.

After the above experiments, we will now select a continuous data sequence of length 4 and a periodic data sequence of length 9 as the inputs of Model-A, Model-B, Model-AC, and Model-BC. In this paper, these two kinds of data are combined into data with dimension 13 as the input of Model-A. At the same time, we add meteorological data to form the input of the other three models as required and then verify and evaluate each model.

As shown in Table 5, the worst is Model-A, with an nRMSE of 0.2209 and the best is Model-BC, with an nRMSE of 0.1743. Certainly, Model-A, Model-B, Model-AC, and Model-BC are all better than just considering continuous data or just periodicity data. After the above discussion, we finally choose Model-BC as the LSTM-FC model. Next, we will describe the parameters of the model that is finalized.

TABLE 7: Evaluation of each model under different input conditions.

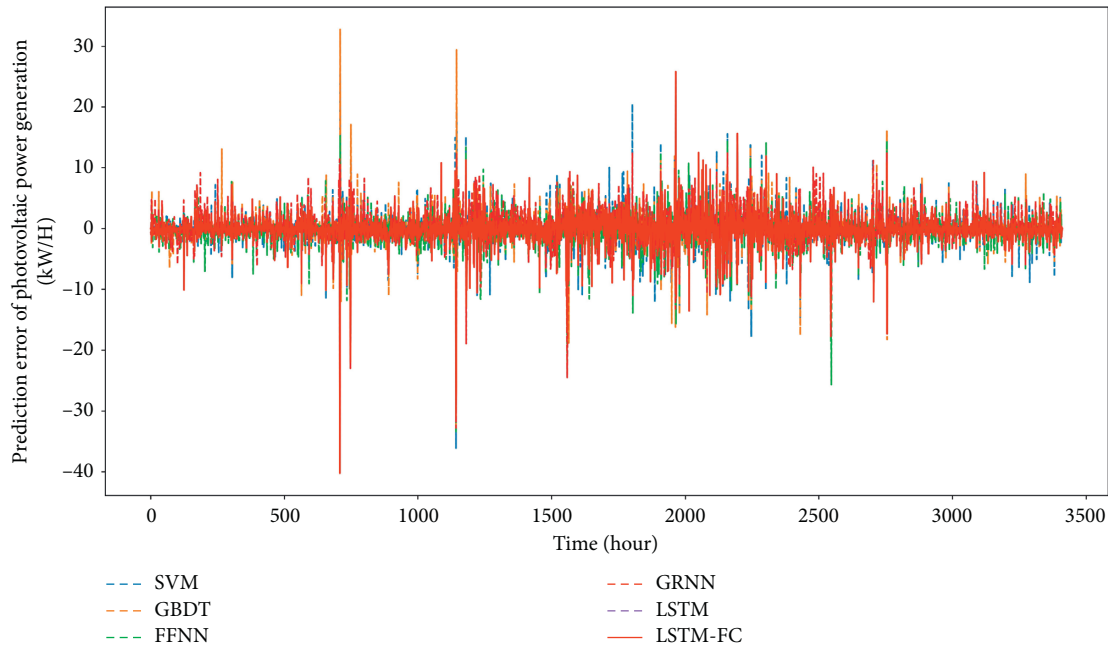|	nRMSE	RMSE	R^2
SVM ($M + C$)	0.2465	3.6200	0.9201
SVM ($M + C + P$)	0.1976	2.9028	0.9486
GBDT ($M + C$)	0.2168	3.1848	0.9381
GBDT ($M + C + P$)	0.1890	2.7762	0.9529
FFNN ($M + C$)	0.2126	3.1223	0.9405
FFNN ($M + C + P$)	0.1918	2.8168	0.9516
GRNN ($M + C$)	0.2433	3.5730	0.9221
GRNN ($M + C + P$)	0.1966	2.8877	0.9491
LSTM ($M + C$)	0.2471	3.6298	0.9196
LSTM ($M + C + P$)	0.1919	2.8189	0.9515
LSTM-FC	0.1743	2.5605	0.9600

FIGURE 10: The error results of different models on test set.

4.3. Network Parameter Setting. The performance of the model will be affected by some factors. For example, if the number of network layers is too small, the accuracy may not be high; if the number of network layers is too large, the result of overfitting may also be caused. Therefore, an optimal parameter should be selected after several times of training. After feedback of training results for many times, the final network parameters in this paper are shown in Table 6, mainly including 2 LSTM layers and 4 FC layers, and Adam optimization function, and the continuous variables and periodic variables determined by the above tests are 4 and 9, respectively.

4.4. Result Analysis. After the above experimental results, we know that Model-BC has the highest accuracy. Following, we will use LSTM-FC to represent Model-BC. We will evaluate LSTM-FC on the test set, and the prediction results obtained by inputting the test set into our model are shown in Figure 9, where the solid line blue is the truth value and

orange is the predicted value. As can be seen from the local enlarged figure, the expected value and the predicted value are basically fitted, and the error between them is very small.

As shown in Table 7, ($M + C$) means to take continuous data and meteorological data as input, and ($M + C + P$) is to add one periodic data as input. We use SVM, GBDT, FFNN, GRNN, and LSTM models commonly used in related work to test different inputs, and the results showed that the model error with periodic data as input was less than the model error not applicable to periodic data. This verifies the idea that using periodic data as input in this paper can improve the accuracy, and under the same, the error of the LSTM-FC model proposed in this paper is minimal. The following model results take ($M + C + P$) as input. Compared with SVM, GBDT, FFNN, GRNN, and LSTM, LSTM-FC's RMSE improves 11.79%, 7.77%, 9.1%, 11.33%, and 9.16%, respectively. Figure 10 shows the errors of each model on the test set, and it can be seen that the errors of LSTM-FC are smaller than other models in most of samples. Through the above discussion, it can be concluded that the performance

of LSTM-FC model in PV power generation prediction is superior to other models.

5. Conclusions

This paper first expounds the benefits of accurate PV power quantity prediction; that is, it can improve the operation efficiency of PV power station, generate stable and reliable power supply, et al. Then, we discuss the research of some current machine learning and deep learning methods in PV power generation prediction. In most of the previous PV power prediction studies, the influence of time continuity and meteorological data on PV power was considered, but did not obtain the periodicity dependence or the model used is not as good as LSTM in capturing time correlation. Therefore, a PV power generation prediction method of LSTM-FC with double branches is proposed. It used LSTMs to better obtain the temporal correlation of PV power generation, used FCs to capture the mapping relationship between meteorological data and PV power generation, and then weighted the output of the two branches to get the final prediction result. The simulation results show that the accuracy of $(M + C + P)$ as input is higher than $(M + C)$, and the accuracy of LSTM-FC model is the highest compared with other baseline models in the same situation, with nRMSE reaching 0.1743 and RMSE reaching 2.5605.

Using the results of this research, we can use the predicted value of PV power generation to provide decision-making for power dispatch, which greatly improves the output efficiency, basically solving the problem of power output reliability caused by the volatility and intermittentness of PV power generation. On the other hand, we can set an outliers confidence interval. When the error between the actual power generation and the predicted power generation exceeds this interval, it indicates that the power station may have a certain fault, which provides feasibility for fault detection. In subsequent studies, we will consider solar irradiance, average power, panel temperature, and rainfall as input features to obtain a better model. Moreover, since the training time of deep learning model is slightly longer than other machine learning models' training time, we will consider how to further reduce the time consumption and make the training faster in the future.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

Supplementary Materials

Here, the authors will explain some of the details of the data set. Considering the confidentiality of the data, the authors collected open photovoltaic power generation data and meteorological data from January 1, 2018, to December 31, 2018, resulting in 13,792 training data and 3,412 test data. As shown in Table 8, the necessary properties mainly include meteorological data, historical continuous data, and historical periodic data. The meteorological data include temperature, humidity, weather, wind direction, and speed. (*Supplementary Materials*)

References

[1] G. K. Singh, "Solar power generation by PV (photovoltaic) technology: a review," *Energy*, vol. 53, pp. 1–13, 2013.

[2] X. Xu, K. Vignarooban, B. Xu, K. Hsu, and A. M. Kannan, "Prospects and problems of concentrating solar power technologies for power generation in the desert regions," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1106–1131, 2016.

[3] IRENA, *Renewable Capacity Statistics 2020*, International Renewable Energy Agency (IRENA), Abu Dhabi, UAE, 2020, https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Mar/IRENA_RE_Capacity_Statistics_2020.pdf.

[4] U. K. Kannan, K. S. Tey, M. Seyedmahmoudian et al., "Forecasting of photovoltaic power generation and model optimization: a review," *Renewable and Sustainable Energy Reviews*, vol. 81, no. 1, pp. 912–928, 2018.

[5] Y. Zheng, X. Yi, M. Li et al., "Forecasting fine-grained air quality based on big data," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15)*, pp. 2267–2276, Association for Computing Machinery, New York, NY, USA, August 2015.

[6] J. Zhang, Y. Zheng, D. Qi et al., "Deep spatio-temporal residual networks for citywide crowd flows prediction," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI'17)*, pp. 1655–1661, AAAI Press, San Francisco, CA, USA, February 2017.

[7] Y. Chen, A. A. Thatte, and L. Xie, "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation," *IEEE Transactions on Sustainable Energy*, vol. 6, pp. 104–112, 2015.

[8] J. Wang, Z. Song, and R. Ran, "Short-term photovoltaic power generation rolling forecast based on optimized SVM," *IN Proceedings of the CSU-EPSA*, vol. 28, no. 11, pp. 9–13, 2016.

[9] A. Lahouar, A. Mejri, and J. Ben Hadj Slama, "Importance based selection method for day-ahead photovoltaic power forecast using random forests," in *Proceedings of the 2017 International Conference on Green Energy Conversion Systems (GECS)*, pp. 1–7, Hammamet, Tunisia, March 2017.

[10] J. Shi, W.-J. Lee, Y. Liu, Y. Yang, and P. Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, vol. 48, no. 3, pp. 1064–1069, 2012.

[11] J. Wang, P. Li, R. Ran, Y. Che, and Y. Zhou, "A short-term photovoltaic power prediction model based on the gradient boost decision tree," *Applied Sciences*, vol. 8, no. 5, p. 689, 2018.

[12] L. Du, L. Zhang, X. Tian et al., "Short-term photovoltaic power forecasting using deep convolutional networks," in *Proceedings of the 2019 2nd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 149–153, Chengdu, China, May 2019.

[13] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, no. 2, pp. 1–22, 2019.

[14] F. Liu, R. Li, Y. Li, R. Yan, and T. Saha, "Takagi-Sugeno fuzzy model-based approach considering multiple weather factors for the photovoltaic power short-term forecasting," *IET Renewable Power Generation*, vol. 11, no. 10, pp. 1281–1287, 2017.

[15] P. Ramsami and V. Oree, "A hybrid method for forecasting the energy output of photovoltaic systems," *Energy Conversion and Management*, vol. 95, pp. 406–413, 2015.

[16] K. Cheng, Q. Sun, and X. Y. Ma, "Power forecast research of photovoltaic system based on double-level neural network," in *Proceedings of the 2018 International Conference on New Energy and Future Energy System*, p. 012007, Shanghai, China, August 2018.

[17] X. Huang, C. Zhang, Q. Li, Y. Tai, B. Gao, and J. Shi, "A comparison of hour-ahead solar irradiance forecasting models based on LSTM network," *Mathematical Problems in Engineering*, vol. 2020, Article ID 4251517, 15 pages, 2020.

[18] D. A. R. de Jesús, P. Mandal, S. Chakraborty, and T. Senjyu, "Solar PV power prediction using a new approach based on hybrid deep neural network," in *Proceedings of the 2019 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, Atlanta, GA, USA, August 2019.

[19] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2727–2740, 2019.

[20] W. Huang, M. J. Bah, and M. Hammad, "Progress in outlier detection techniques: a survey," *IEEE Access*, vol. 7, pp. 107964–108000, 2019.