

Research Article

Short-Term Prediction Method of Solar Photovoltaic Power Generation Based on Machine Learning in Smart Grid

Yuanyuan Liu 

Department of Mathematics, Lyuliang University, Luliang 033001, Shanxi, China

Correspondence should be addressed to Yuanyuan Liu; 20181029@llu.edu.cn

Received 1 August 2022; Revised 21 August 2022; Accepted 26 August 2022; Published 12 September 2022

Academic Editor: Baiyuan Ding

Copyright © 2022 Yuanyuan Liu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to improve the accuracy of ultra short-term power prediction of the photovoltaic power generation system, a short-term photovoltaic power prediction method based on an adaptive k-means and Gru machine learning model is proposed. This method first introduces the construction process of the model and then builds a short-term photovoltaic power generation prediction model based on an adaptive k-means and Gru machine learning models. Then, the network structure and key parameters are determined through experiments, and the initial training set of the prediction model is selected according to the short-term photovoltaic power generation characteristics. And the adaptive k-means is used to cluster the initial training set and the photovoltaic power on the forecast day. The Gru model is trained on the initial training set data of each category, and the generated power is predicted in combination with the trained Gru model. Finally, considering three typical weather types, the proposed method is used for simulation analysis and compared with the other three traditional photovoltaic power generation single prediction models. The comparison results show that the proposed short-term photovoltaic power generation prediction method based on an adaptive k-means and Gru network has better effect, better robustness, and less error.

1. Introduction

In the 21st century, with the increasingly fierce global economic competition, the demand for energy in various countries is increasing, and the energy problem has become a key factor affecting the international influence of countries [1]. However, with the continuous growth of social and economic development demand, the use and consumption of traditional fossil energy are also increasing, which not only causes the global shortage of traditional fossil energy but also causes a lot of pollution to the environment, and these pollution are serious and difficult to control. There is no doubt that this will run counter to China's proposal to build a resource-saving and environment-friendly society [2–4].

There are many renewable energy sources on Earth, including solar energy, wind energy, and many other energy sources, but all renewable energy sources have a common feature, that is, it is not easy to collect, and there is great inconvenience. Compared with other energy sources, solar

energy is easy to obtain, the use conditions are unlimited, and there are no major technical barriers to its application. At the same time, it is inexhaustible. It is a new type of green energy widely used in the world at present. The most direct use of solar energy is photovoltaic power generation [5–7].

As the largest developing country, China's annual total solar radiation is between 928 kWh/m^2 – 2333 kWh/m^2 , and the annual average solar radiation is 1626 kWh/m^2 . In most regions of China, the daily average radiation can reach 4 kWh/m^2 , which shows that China has superior natural conditions for photovoltaic power generation [8–11]. Globally, the newly added PV installed capacity increased from 39.6 gw in 2010 to 580.16 gw in 2019. The specific growth of each year is shown in Figure 1.

With the policy support of encouraging the development of photovoltaic power generation in various aspects, China has made breakthroughs in photovoltaic power generation technology in recent years. Especially for some central and western regions of China, affected by other energy structures, they have to transform in this direction. At the same

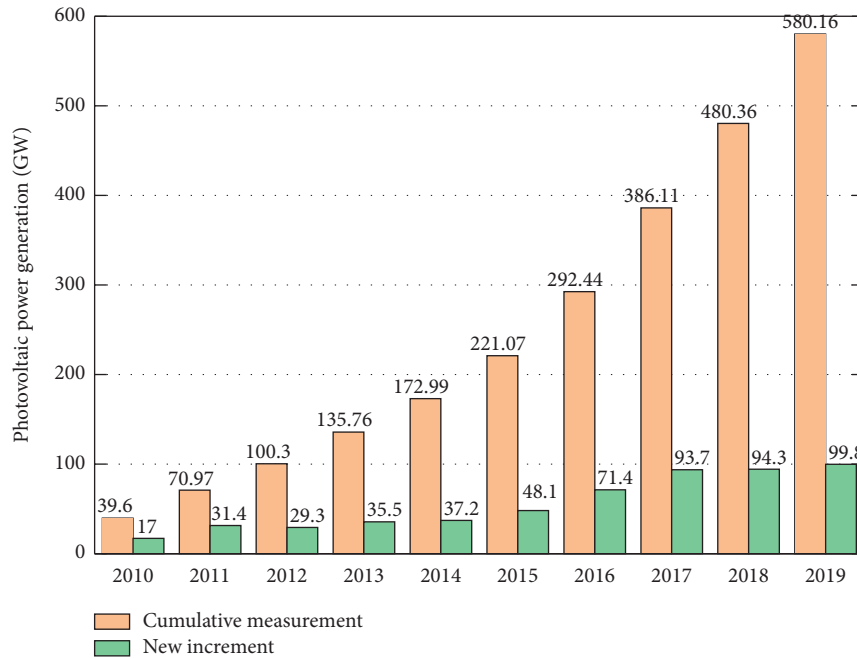


FIGURE 1: Global cumulative and newly added PV installed capacity (2010–2019).

time, more and more difficulties are faced, and the research on the prediction of photovoltaic power generation is also increasing year by year.

Photovoltaic power generation is the most important way for humans to use solar energy at present. It will not affect the environment during this utilization process but it has the advantages of short construction period, mature technology, large-scale development, and sustainable development. It has broad development prospects and is more and more valued by people [12]. Since photovoltaic power generation can only be carried out under sunshine, and there are weather changes such as day and night alternation and cloudy, sunny, rainy, and snowy on Earth, the photovoltaic power station can only generate electric energy during the day. It is a typical intermittent power supply. Its power generation is affected by meteorological conditions such as solar irradiation intensity and ambient temperature and has great volatility and randomness [13–16].

These characteristics of photovoltaic power generation will have a severe impact on the stable operation of the power grid when large-scale photovoltaic power generation is connected to the grid and have a negative impact on the entire power system, which is one of the important challenges faced by large-scale photovoltaic power generation. Relevant research shows that when the grid-connected capacity of photovoltaic power generation accounts for more than 15% of the total power generation of the power system, its fluctuation may cause the paralysis of the power system [3]. If the photovoltaic power generation power can be predicted timely and accurately, the impact of the fluctuation of photovoltaic power generation on the power grid will be greatly reduced, which is of great significance to the power grid dispatching and the operation of photovoltaic power stations [17, 18]. It is an important topic for researchers in relevant fields at home and abroad to use some

energy transfer equations or mathematical statistical algorithms to predict the photovoltaic power generation power in the future, know the output of photovoltaic power stations in advance and provide a reference for power grid dispatching and power station operation [19–24]. However, errors will inevitably occur in the prediction of photovoltaic power generation. If the error of the prediction method used is too large, it will pose a serious threat to the reliability of the photovoltaic output prediction system developed and cannot be popularized. As the proportion of photovoltaic power generation capacity in the total power generation capacity of the power system increases, the impact of these prediction errors on the stability of the power system also increases. Therefore, in order to increase the proportion of photovoltaic power generation capacity in the power system and improve the scale of photovoltaic power generation to expand its economic and social benefits, it is of great significance to make a short-term prediction of photovoltaic power generation [25–27].

The existing short-term photovoltaic power prediction methods can be roughly divided into two kinds, one is the physical method, the other is the statistical method. The physical method is generally to establish the prediction model through the solar radiation transfer equation and the operation equation of photovoltaic equipment; Statistical methods use the relationship between historical operation data to establish prediction models [28]. Because the physical prediction model needs to know the specific parameters of solar radiation equation and photovoltaic modules, its generalization ability is not strong, and the prediction accuracy is significantly lower than that of statistical methods, which have been less used. With the continuous breakthrough of human beings in the process of intellectualization and the continuous investment in research, scholars from various countries began to extend the

research on photovoltaic power prediction methods from traditional statistical methods to artificial intelligence algorithms and achieved good results in photovoltaic power prediction research based on the neural network. Now the process of intellectualization has been deeply rooted in the hearts of Chinese and foreign people. No matter in any field or anywhere, people expect to realize intellectualization as soon as possible. This is also an important field of current research, and its performance is more superior and advanced than other methods. However, the traditional ANN is easy to fall into a trap called local extremum during training, and its accuracy still has great room for improvement. Moreover, the existing high-precision ANN prediction methods need high-precision input data, and it needs to manually select the factors that have a great impact on the photovoltaic power as the input samples, and a lot of preprocessing work needs to be performed on the sample data before training the neural network. When the sample is complex and the feature density is low, the shallow Ann may not be able to effectively learn the internal relationship between the input data and the output data, resulting in low prediction accuracy [29].

By comprehensively analyzing the research status of the above two different directions of point prediction and interval prediction, it can be found that the research and prediction with the increasing demand for photovoltaic in the world, and the research results emerge in endlessly. Among them, the model established by using the deep learning method has high prediction ability, and it is also one of the main directions of photovoltaic generation point prediction in the future.

2. Theoretical Basis of Photovoltaic Power Prediction

2.1. Adaptive k-Means Theory. In the case of three typical weather types, clustering the photovoltaic output power, respectively, can build a more targeted prediction model and further improve the prediction accuracy. Traditional k-means cannot actively determine the number of clusters without knowing the data set, so it is not suitable to cluster the input data set directly. Therefore, this paper introduces an adaptive k-means, which can automatically set the number of clusters according to the input data set. Its main idea is an iterative process based on distance [30]. The steps of k-means algorithm are as follows:

- (1) First, according to our prediction needs, we build targeted models, mainly to determine the number of inputs, outputs, and interneurons
- (2) The second step is to screen the constructed data set and randomly select k data as a center of our initial blood drug clustering, which is recorded as: $\{\lambda_1, \lambda_2, \dots, \lambda_k\}$
- (3) The third step is to calculate the Euclidean distance between the remaining samples and the cluster center through the algorithm and assign it to the nearest cluster center to form K clusters. The

distance measurement formula is given in the following formula:

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}. \quad (1)$$

- (4) In the fourth step, the cluster center is updated by the distance measurement method to be the mean value of all the samples belonging to the cluster
- (5) After the above four steps are completed, a basic unit step is completed. After that, only steps (1), (2), and (3) need to be repeated until the algorithm converges

In addition, in order to reduce some unnecessary errors caused by manual operation to the calculation of the model, we have also made appropriate improvements to this method, mainly through some optimization methods of quantitative search samples to realize the automatic optimization of clustering. K-means selected in this paper is one of these methods, and its specific definition is shown in the following formula:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{C_i + C_j}{D_{i,j}} \right). \quad (2)$$

In order to avoid generating too many clusters, a threshold is used to limit the number of clusters, which is recorded as k_{\max} . The adaptive k-means clustering process is shown in Figure 2.

The strong randomness and large fluctuation of photovoltaic power generation are largely related to meteorological and environmental factors. Temperature, humidity, wind speed, total radiation, air pressure, and other factors have varying degrees of influence on photovoltaic power generation. Therefore, in this process, it is important to choose the right method, and this method can correctly express our impact on the short-term prediction of photovoltaic power generation. Through the expression of reference 28, we can find that this kind of influence factors are numerous and vary greatly and are related to various factors such as regions, among which the biggest influence factor is climate, which is mainly reflected in the variable factors of sunshine in the region. If the sunshine is sufficient and long, the photovoltaic power generation efficiency is high, otherwise the opposite is true. Therefore, when we choose variables and determine factors, we mainly rely on the above views.

2.2. GRU. GRU is a new neural network model developed from the deficiency of LSTM. The proposal of LSTM solves the problem of the RNN model. There is also a gated cyclic unit network in the cyclic neural network, which also solves the problem that RNN cannot deal with the dependence of large time step distance. GRU is a variant of LSTM and can also well capture long-distance dependency problems. The difference between GRU and LSTM lies in the number of gate units. GRU network simplifies the three gates of LSTM into gate units: the update gate and the reset gate. The reset

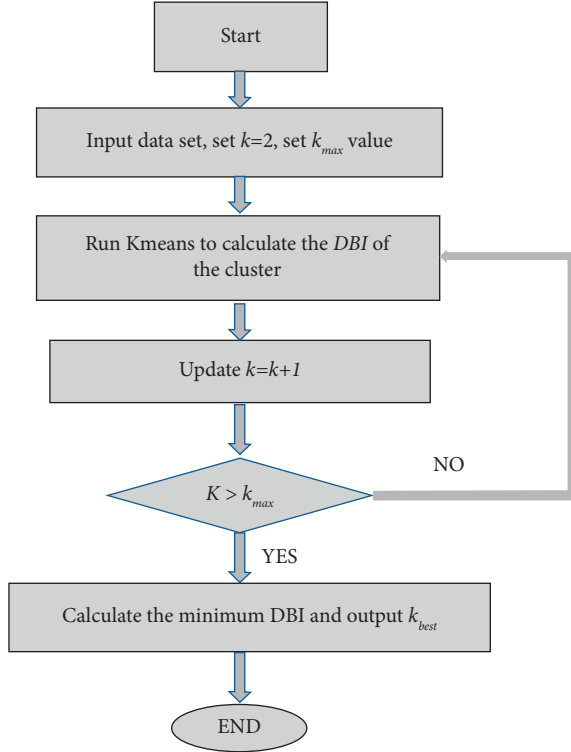


FIGURE 2: Adaptive k-means clustering flow chart.

gate is a combination of a memory cell and a hidden layer. Its function is to control the transmission of the hidden state information of the previous time to the candidate hidden state of the current time and then reset the candidate hidden state information at the current time. The update gate is a combination of the forgetting gate and the input gate. Its function is to realize the hidden function that the original structure does not have through its own designed structure and to update this hidden information through its own special structure. By simplifying the gate structure, the network model parameters are reduced, the operation becomes simpler, and the performance of the network model is improved. GRU network structure is shown in Figure 3.

As shown in Figure 3, GRU network is divided into reset gate r_t , update gate z_t , \tilde{H}_t , and H_t . The calculation expressions for updating doors, resetting doors, and hidden states are as follows.

2.2.1. Update Gate. The input is the input X_t at the current time and the \tilde{H}_{t-1} at the back time, and the output of the update gate is z_t . z_t is the linear combination of X_t and \tilde{H}_{t-1} , and then input to Sigmoid function to get a value from 0 to 1.

$$z_t = \sigma(W_z X_t + U_z H_{t-1} + b_z). \quad (3)$$

2.2.2. Reset Gate. The input is still the input X_t at the H_{t-1} at the back time, and the output of the reset gate is r_t . r_t is the linear combination of X_t and H_{t-1} , and then input to Sigmoid function to get a value from 0 to 1.

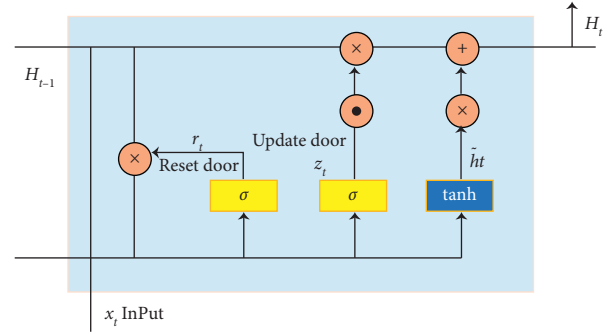


FIGURE 3: GRU network structure.

$$r_t = \sigma(W_r X_t + U_r H_{t-1}). \quad (4)$$

2.2.3. Candidate Hidden State. Multiply the reset gate output r_t and the H_{t-1} at the previous time by elements, and input the operation result into the tanh function after linear combination with the current input X_t , so that the value of the \tilde{H}_t is between -1 and 1 .

$$\tilde{H}_t = \tanh(W X_t + U(r_t \odot H_{t-1}) + b_h). \quad (5)$$

From the above formula, we can see the function of reset gate r_t . When $r_t = 0$, the result of element multiplication between the reset gate output r_t and the previous hidden state H_{t-1} is 0. It means that the H_{t-1} has no effect on the \tilde{H}_t , which is equivalent to discarding the H_{t-1} information. Then the H_t at the current time is only related to the input X_t at the current time, which resets the hidden state. Because of this, resetting the gate at this time helps to capture short-term dependencies.

2.2.4. Hidden State. The update gate output z_t is linearly combined with the H_{t-1} at the previous time and the candidate hidden state at the current time.

$$H_t = (1 - z_t)\tilde{H}_t + z_t \odot H_{t-1}. \quad (6)$$

From the above formula, we can see the function of the update gate z_t . When $z_t = 1$, then $1 - z_t = 0$. At this time, the hidden state H_{t-1} of the previous time completely gives the hidden state H_t of the current time and completely retains the hidden state of the current time. At this time, if there is a long-term dependency, the hidden state information can also be transmitted and retained. Because of this, GRU can capture long-term dependencies, which is also the most critical part of the GRU network. Then updating doors helps capture long-term dependencies.

3. Construction of the Joint Prediction Framework Based on Adaptive K-Means and GRU

3.1. Construction of the Joint Prediction Model. Because the photovoltaic output power curve has obvious similarity characteristics under similar weather conditions, this paper

forecasts the photovoltaic output power under three typical weather conditions. The main steps of model construction are as follows:

- Step 1: according to the input and output of the model, determine the basic input and output units of the model, as well as the number of neurons and the type of activation function;
- Step 2: considering the weather conditions of the day to be predicted, select the historical similar day data as the initial training set;
- Step 3: normalize all data to obtain the training set and the test set;
- Step 4: use adaptive k-means for clustering analysis of historical data and predicted daily data under each weather and combine Gru for training and prediction;
- Step 5: get the results of the photovoltaic power generation prediction model through model training.

The specific process of the model constructed by this process is shown in Figure 4.

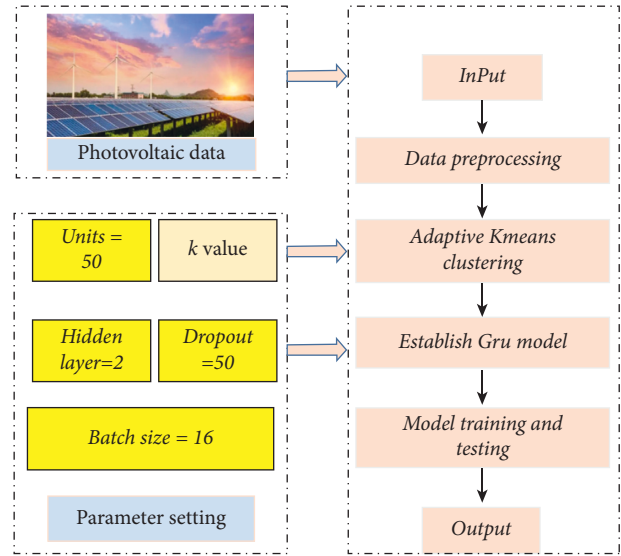


FIGURE 4: Prediction framework of photovoltaic output power based on adaptive k-means and GRU.

3.2. Establishment of the Photovoltaic Power Generation Prediction Data Set

3.2.1. *Introduction to Original Data.* The experimental data set in this paper is a photovoltaic power generation data set in a certain region. This data set records the relevant power generation data of more than 100 users equipped with solar power generation devices in 2020, and the data sampling frequency is 1 hour. That is, taking the power station as the unit, the respective power and meteorological data are recorded. The fields used in the power generation data table are shown in Table 1. The relationship between the three is: use = gen + grid. Use is the total power consumption of each part of the photovoltaic power station; Gen refers to the power generated by solar photovoltaic power generation itself; Grid refers to the power that the power station is connected to the power grid. A positive value indicates that it is connected to the power grid, and a negative value indicates that it is output to the power grid. The other two powers are always positive.

The format of the above power generation data set is different due to the different types of sensors configured by each household, and the format of the power consumption field is also different. When the time information is recorded, it mainly includes some data sets in the following three formats: type I data set contains a Gen (generated power) field, which can be directly used in the prediction experiment of photovoltaic output; Type II does not contain a Gen field, which needs to be obtained by subtracting grid data from use field; The third type is automatic accumulation during recording. To obtain the generated power of this period, you can obtain it by subtracting the data of adjacent periods. The unit of the three is kW, so the unit of photovoltaic output power in the full text is also kW.

The weather data matches the power generation data, and its specific description is shown in Table 2.

TABLE 1: Field description of photovoltaic power generation data.

Category	Full name
Date & time	Time
Use (kW)	Power consumption
Gen (kW)	Generating power
Grid (kW)	Grid power

Figure 5 shows the photovoltaic output power in one week. It can be seen that photovoltaic power generation has obvious diurnal periodicity, volatility, and randomness. Therefore, we will rely on the powerful feature learning ability of the deep learning algorithm to analyze and predict it and realize the effective capture of its characteristics.

3.2.2. *Data Preprocessing.* It can be seen from Figure 5 that the power generation data generally has the characteristics of stable change trend and relatively concentrated values. However, in some special times and stages, a part of the data is usually incomplete (lacking some interesting attribute values), inconsistent (including code or name differences), and vulnerable to noise (errors or abnormal values). At the same time, when the database is too large, the data sets often come from multiple heterogeneous data sources, showing low-quality data, and the results obtained from network training are often poor. Therefore, it is necessary to preprocess the original power generation data to eliminate the influence of the dimension of the data itself and other useless features to ensure the effectiveness of the sample data training, and it is necessary to carry out standardized processing.

At present, the normalization of maximum and minimum values based on numerical linear transformation is often adopted at home and abroad, which restricts the

TABLE 2: Description of meteorological data.

Category	Temp	Dewpt	Hum	Wgust	Wdird	Wdire
Description	Temperature	Dew time	Humidity	Wind speed	Wind direction azimuth angle	Wind direction
Category	Vis	Press	Windchill	Heatindex	Precipm	Conds
Description	Visibility	Pressure	Wind-chill index	Calorific value	Precipitation	Weather type

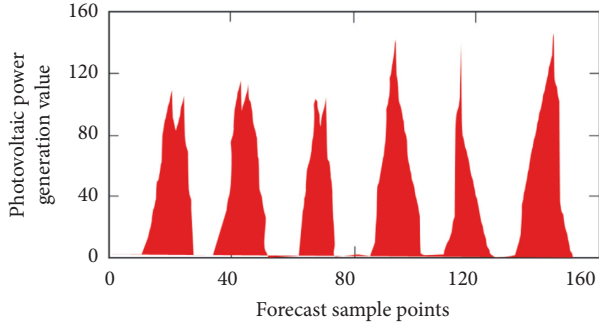


FIGURE 5: Actual output power of photovoltaic power generation.

TABLE 3: Description of meteorological data.

Evaluation value	Evaluation results
<15%	Excellent accuracy
16%–25%	Good accuracy
25%–50%	Average accuracy
>50%	Unreasonable accuracy

original data values to $[0,1]$, that is, normalizing the data through the following formula:

$$X^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}}, \quad (7)$$

where X_{\max} and X_{\min} and X^* , respectively, represent the maximum and minimum values of photovoltaic data types of the overall data set, as well as the standardized data obtained from the standardization of this formula, whose size is not greater than 1 and not less than 0.

3.3. Selection of Prediction and Evaluation Indicators for Photovoltaic Power Generation. The evaluation methods of the prediction algorithm mainly include: root mean square error RMSE, average absolute percentage error MAPE, square sum error SSE, mean square error MSE, and average absolute error MAE. In this paper, the root mean square error RMSE and average absolute percentage error MAPE are used to evaluate the prediction results.

- (1) The formula of root mean square error RMSE is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (P'_f - P'_a)^2}{N}}, \quad (8)$$

where N is the predicted quantity; P'_f is the prediction data; P'_a is the actual data; i is the prediction time.

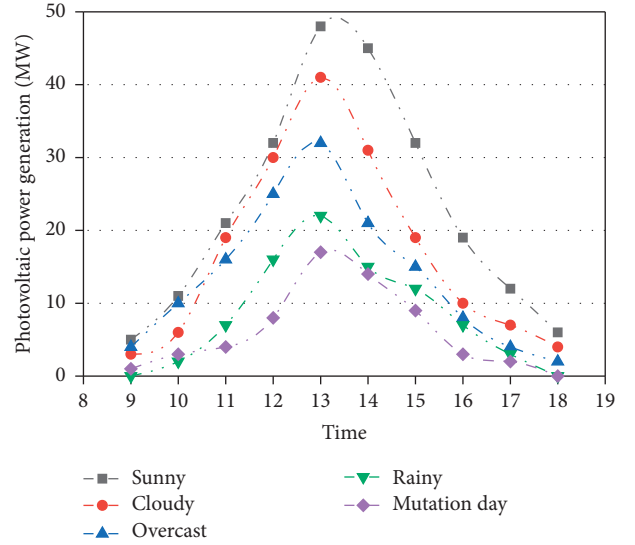


FIGURE 6: Photovoltaic power generation under different weather conditions.

- (2) The average absolute percentage error MAPE, whose formula is as follows:

$$\text{MAPE} = \frac{100}{N} \sum_{i=1}^N \left| \frac{P'_f - P'_a}{P'_a} \right|_0\%, \quad (9)$$

where N is the total number of sample data; P'_f is the i th predicted value; P'_a is the i th actual value. The MAPE evaluation criteria for evaluating the prediction accuracy are shown in Table 3.

3.4. Model Input Feature Selection. This paper selects the photovoltaic power data in a certain interval of a photovoltaic power generation system and collects 24 sample points every day. It can only select the period of stable output of photovoltaic power for analysis. The photovoltaic power generation power under different weather is shown in Figure 6. When the weather is relatively stable, the photovoltaic power generation power is the highest in sunny weather, and the others are cloudy, cloudy and rainy, and snowy weather in turn. In stable weather, the power fluctuation of the photovoltaic system is small, and the output is relatively stable, which is close to the parabola as a whole. In sudden change weather, photovoltaic power generation fluctuates greatly, which has a great impact on the stable operation of the entire power grid. Therefore, distinguishing

weather types is of great significance for the prediction of photovoltaic power generation.

To sum up, we can know that there are many factors that can affect photovoltaic power, but it mainly includes weather and time period. Therefore, the characteristic input of photovoltaic power prediction in this paper mainly includes the above two kinds.

3.5. Prediction Steps of Photovoltaic Short-Term Power Generation. The combined neural network prediction model established in this paper mainly highlights the impact of day type on the output power of the photovoltaic power generation system. When establishing the model, we need to focus on this point and improve the weight of day type index and do not need to consider the impact of weather factors when predicting. It can adapt to all kinds of weather and has strong adaptability and good prediction ability.

3.5.1. Determining Input Data. The number of nodes in the input layer corresponds to the input variables of the prediction model. The prediction model in this paper has a total of 12 input variables. Considering the specific geographical location of the photovoltaic power station selected in the paper, combined with historical data, the power generation between 18:00 a.m. in a day and 8:00 a.m. in the next day is almost 0, so the power generation of 10 power generation time series from 9:00 a.m. to 18:00 a.m. on the day before the prediction day is selected as the input to the prediction model.

In the prediction of photovoltaic power generation output, it is usually necessary to divide the sub models skillfully according to the type of day, otherwise the prediction model may fail. According to the factors that affect the photovoltaic power generation system, because the type of day has a large influence factor in the influencing factors of photovoltaic power generation output, the paper further increases the weight of the type of day index when establishing the prediction model.

3.5.2. Determining the Output Layer. The prediction result is to predict the generation output power in each period of the day, so there are n nodes at the output end, corresponding to the hourly output power between 9:00 and 18:00, respectively.

4. Case Analysis

4.1. Sample Data Selection. Considering the weather type, the photovoltaic power in sunny, cloudy, and rainy days is predicted by using the prediction model in this paper. The data comes from the field measured data of a city in China, including total radiation, humidity, temperature, and photovoltaic power generation. The data is the real data of 2020. The selected time period is between 9:00 and 18:00, and the sampling interval is 15 minutes, that is, the number of sampling points in a day is 40. The predicted days are the first period in the period, i.e., sunny days. The second period was

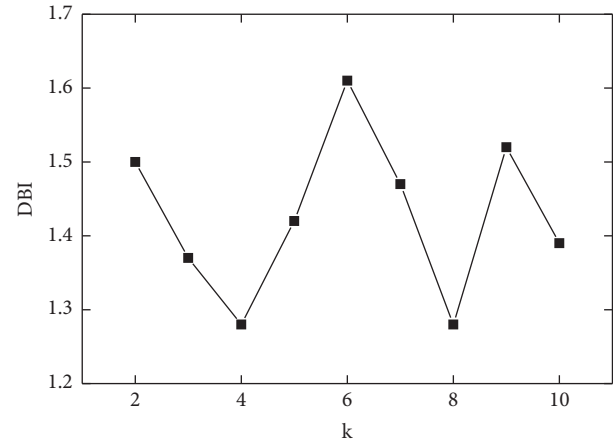


FIGURE 7: DBI with different k values.

cloudy and the third period was rainy. Due to the lack of snow data, it is not included in the analysis and prediction. The power fluctuation caused by meteorological factors in adjacent periods is small. Therefore, the data of similar days with the same weather similar to the forecast day is selected as the initial training set. According to the type of forecast days, the similar 10 days are selected as the initial training set.

4.2. Cluster Analysis. Taking the total radiation, humidity, and temperature of similar days and forecast days as inputs, the photovoltaic power generation power under three weather types is clustered, respectively. The initial training set is 400 time sampling points, plus the data of the forecast day, a total of 520 data. According to the previous analysis, adaptive k -means are used for clustering. Considering the length of the data, k_{\max} is set to 10. Take time period 1 (sunny) as an example for analysis. The value range of K is [2, 10]. DBI when k takes different values is shown in Figure 7.

According to Figure 7, when k is taken as 6, DBI is the largest, which is 1.61; When k takes 4 and 8, DBI is the smallest, which is 1.28. At this time, the clustering effect is the best. Therefore, the data of the initial training set and the forecast day on sunny days are divided into three categories, which are called class 1, class 2, and class 3 here. Similarly, adaptive k -means are used to cluster the data of the initial training set and the predicted day in cloudy and rainy days, respectively. When k is taken as 4 when it is cloudy, DBI is the smallest, which is 1.12. In rainy days, K is taken as 4, and DBI is the smallest, which is 1.46.

4.3. Analysis of Prediction Results. In order to compare the advantages of the model constructed in this paper, in addition to the training of the constructed model, this paper also selects two other commonly used short-term prediction models (LSTM and BP) and a single GRU model for comparative analysis.

Among them, the photovoltaic power prediction results in sunny weather are shown in Figure 8, and the prediction

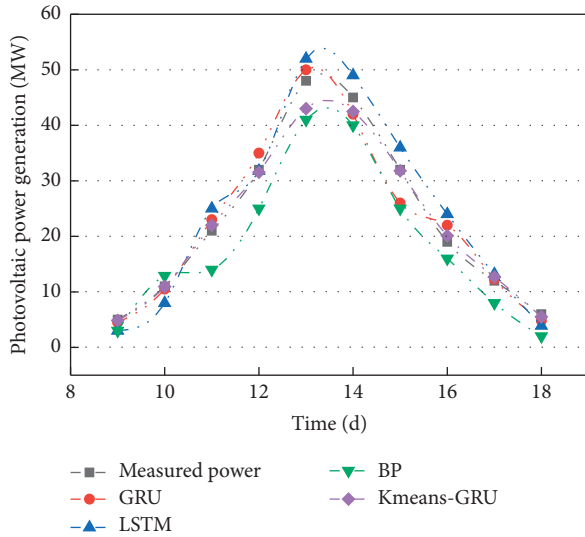


FIGURE 8: Photovoltaic power prediction results of different models on sunny days.

TABLE 4: Photovoltaic power prediction error in sunny weather.

Model	RMSE	MAPE/(%)
GRU	16.33	0.13
LSTM	19.24	0.21
BP	27.18	0.37
Kmeans-GRU	8.15	0.04

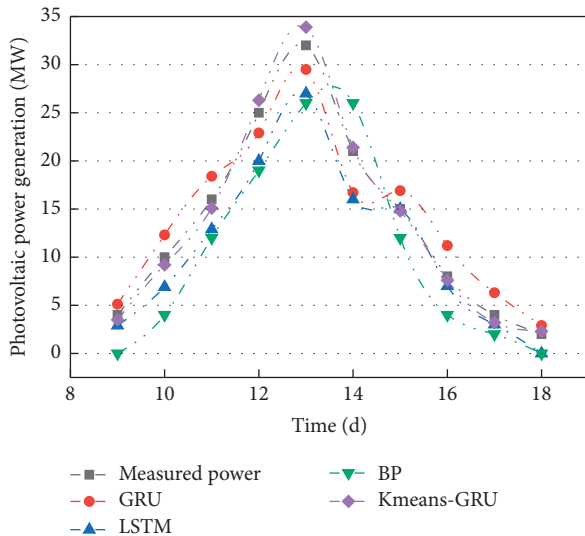


FIGURE 9: Photovoltaic power prediction results of different models for overcast world.

error is shown in Table 4. In sunny, the fluctuation of photovoltaic output curve is small, and the power change has a certain regularity. The four models show good prediction results. By analyzing MAPE and RMSE, it can be found that the error of the proposed adaptive k-means GRU prediction model is smaller than that of the single prediction model and the other two models, and the accuracy of the single model

TABLE 5: Photovoltaic power prediction error in cloudy weather.

Model	RMSE	MAPE/(%)
GRU	19.17	0.17
LSTM	23.16	0.27
BP	35.44	0.42
Kmeans-GRU	10.12	0.08

prediction is slightly reduced when the power curve fluctuates slightly at noon. Compared with GRU, LSTM, and BP model, the proposed prediction model based on adaptive k-means GRU can be better close to the actual power curve as a whole, and the fitting effect is the best.

The photovoltaic output power prediction results and RMSE and MAPE results under cloudy weather conditions are shown in Figure 9 and Table 5, respectively. In cloudy weather, the sunshine is obviously lower than that in sunny weather due to the influence of weather. It is also due to the change of such complex factors that it is difficult for each model to control such variable factors, thus it is easy to lead to a series of prediction errors. The prediction results and RMSE and MAPE results are prone to large deviations, and the comparison of the prediction accuracy of various models is thus revealed. It reduces the influence of power data fluctuation on the accuracy of the model and improves the prediction performance of the model during the period when weather conditions fluctuate. In cloudy weather, the MAPE value of k-means GRU is lower than that of GRU, LSTM, and BP, and the values of each model are 0.04, 0.13, 0.21, and 0.37, respectively. It can be seen that the error value of the k-means GRU model is significantly lower than that of other models and is 10.81% of that of the BP traditional neural network model. Therefore, according to the above, it can be concluded that the prediction performance of the k-means GRU model for photovoltaic is better and the prediction deviation is the lowest.

The prediction results of photovoltaic output power and RMSE and MAPE results under rainy weather conditions are shown in Figure 10 and Table 6. Under the condition of rainy weather, the influence of weather is more obvious than that of overcast weather, and its sunshine is significantly lower than that of overcast weather, accompanied by the interference of rain. Under the change of this more complex factor, the conditions for controlling this variable factor are more stringent, which leads to more prediction errors, and its prediction results and RMSE and MAPE results are prone to large deviations. The comparison of the prediction accuracy of various models is thus easier to show. When weather condition fluctuates from time to time, it is also called sudden change weather, which is found through the prediction results. Under the sudden change weather of rainy days, the MAPE value of k-means GRU is lower than that of GRU, LSTM, and BP, and the values of each model are 0.08, 0.17, 0.27, and 0.42, respectively. In contrast, other models jump out of the predictable range and the acceptable error range. Only the model error constructed in this paper is still within the controllable range, and the intelligent k-means GRU model has a better prediction performance for photovoltaic, with the lowest prediction deviation.

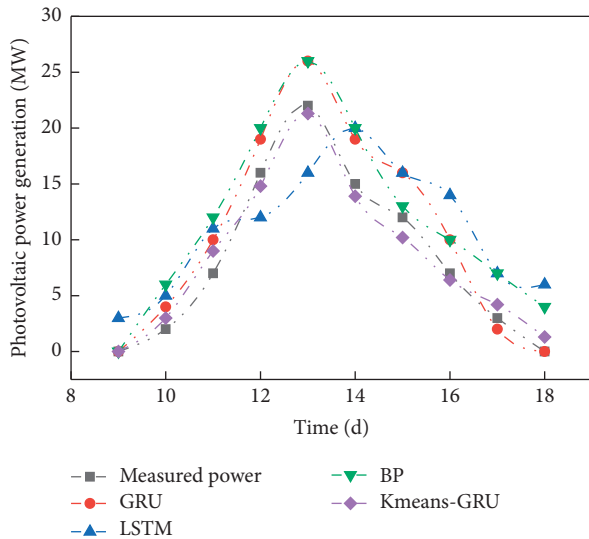


FIGURE 10: Photovoltaic power prediction results of different models in rainy days.

TABLE 6: Photovoltaic power prediction error in rainy days.

Model	RMSE	MAPE/(%)
GRU	23.28	0.21
LSTM	42.36	0.33
BP	55.54	0.57
Kmeans-GRU	14.29	0.16

Based on the short-term prediction results of the above four models for photovoltaic power generation, it can be found that the adaptive k-means Gru model constructed in this paper has better advantages than other models in terms of the prediction of sudden weather, such as sunny, cloudy, or rainy days. Its errors are lower than other models, and its accuracy is higher, which verifies the effectiveness of the adaptive k-means Gru short-term photovoltaic power generation prediction model constructed in this paper.

5. Conclusion

In order to solve the short-term prediction accuracy temperature of photovoltaic power, this paper divides the weather types and proposes a photovoltaic power ultra-short-term prediction model based on the adaptive k-means GRU method. The adaptive k-means is directly used to cluster the initial training set and the photovoltaic power of the prediction day and find out the local characteristics of the data and predict the power. Three single models are established to compare with the proposed model, and the prediction error is evaluated according to MAPE and RMSE. The proposed model solves the problem of low accuracy of traditional prediction methods in power fluctuations. The main conclusions are as follows:

- (1) Photovoltaic output power has great randomness, among which there are many influencing factors, including injection time and weather, especially the

weather type has a great impact on photovoltaic output power.

- (2) By using the combination of adaptive k-means and Gru models, the data preprocessing and clustering analysis of the initial training set and the photovoltaic power generation on the forecast day can significantly improve the prediction accuracy of the model.
- (3) Through comparison with other models, it is found that the prediction performance and stability of the model proposed in this paper are better in sunny days and cloudy days, and the prediction performance is improved in rainy days. The prediction error value of the model is significantly lower than that of other neural network models, and only 10.81% of that of the BP neural network, thereby indicating the reliability of the model.
- (4) According to the prediction results of sunny, cloudy, and rainy days, the prediction accuracy of the adaptive k-means and Gru combined model meets the ultra-short-term prediction requirements of the output power of the photovoltaic power generation system. The difference between predicted power and actual power is small, and the prediction error value does not affect the normal operation of the system.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that they have no conflicts of interest.

Acknowledgments

The study was supported by the Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi Province, Studies of solar irradiance forecasting using a mesoscale meteorological model (No. 2019L0983); the Introduction of High-level Science and Technology Talent Program of Luliang Development Area in Shanxi Province, and Studies of solar irradiance forecasting model based on Meteorological Big Data (No. 2019107).

References

- [1] P. Sasinee and U. Toshimitsu, "Supervisory control of partially observed quantitative discrete event systems for fixed-initial-credit energy problem[J]," *IEICE - Transactions on Info and Systems*, vol. 100, no. 6, pp. 1166–1171, 2017.
- [2] G. Huey, L. Wang, R. Dave, R. R. Caldwell, and P. J. Steinhardt, "Resolving the cosmological missing energy problem," *Physical Review D - Particles and Fields*, vol. 59, no. 6, 1998.
- [3] A. L. Demain, "Biosolutions to the energy problem," *Journal of Industrial Microbiology & Biotechnology*, vol. 36, no. 3, pp. 319–332, 2009.

- [4] Y. Guo, D. L. Thompson, and T. D. Sewell, "Analysis of the zero-point energy problem in classical trajectory simulations," *The Journal of Chemical Physics*, vol. 104, no. 2, pp. 576–582, 1996.
- [5] S. Y. Lin, J. Moreno, and J. G. Fleming, "Three-dimensional photonic-crystal emitter for thermal photovoltaic power generation," *Applied Physics Letters*, vol. 83, no. 2, pp. 380–382, 2003.
- [6] J. G. . Fleming, "Addendum: "three-dimensional photonic-crystal emitter for thermal photovoltaic power generation" [appl. Phys. Lett. 83, 380 (2003)]," *Applied Physics Letters*, vol. 83, Article ID 249902, 2003.
- [7] M. Hosenuzzaman, N. A. Rahim, J. Selvaraj, M. Hasanuzzaman, A. Malek, and A. Nahar, "Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation," *Renewable and Sustainable Energy Reviews*, vol. 41, no. jan, pp. 284–297, 2015.
- [8] M. Hosenuzzaman, N. A. Rahim, J. Selvaraj, M. Hasanuzzaman, A. A. Malek, and A. Nahar, "Global prospects, progress, policies, and environmental impact of solar photovoltaic power generation," *Renewable and Sustainable Energy Reviews*, vol. 41, 2015.
- [9] R. Li and G. Li, "Photovoltaic power generation output forecasting based on support vector machine regression technique," *Electric power*, vol. 41, 2008.
- [10] Y. Wang, Y. Song, and X. Zong, "Two-level optimal dispatch of power system with photovoltaic power generation based on green certificate trading and generation rights trading," *Journal of China University of Petroleum(Edition of Natural Science)*, 2019.
- [11] X. Zhou, D. Song, Y. Ma, and D. Cheng, "Grid-connected control and simulation of single-phase two-level photovoltaic power generation system based on repetitive control," in *Proceedings of the International Conference on Measuring Technology & Mechatronics Automation*, March 2010.
- [12] K. Furushima and Y. Nawata, "Performance evaluation of photovoltaic power-generation system equipped with a cooling device utilizing siphonage," *Journal of Solar Energy Engineering*, vol. 128, no. 2, pp. 146–151, 2006.
- [13] M. Jamil, S. Kirmani, and M. Rizwan, "Techno-economic feasibility analysis of solar photovoltaic power generation: a review," *Smart Grid and Renewable Energy*, vol. 03, no. 04, pp. 266–274, 2012.
- [14] K. Furushima, Y. Nawata, and M. Sadatomi, "Performance Evaluation of Photovoltaic Power-Generation System Equipped with a Cooling Device Utilizing Siphonage," *2nd Report 改善 of a PV System Placed on a Residential Rooftop*, vol. 128, 2005, <https://doi.org/10.1115/1.2183805>.
- [15] W. Fang, Q. Huang, S. Huang, J. Yang, E. Meng, and Y. Li, "Optimal sizing of utility-scale photovoltaic power generation complementarily operating with hydropower: a case study of the world's largest hydro-photovoltaic plant," *Energy Conversion and Management*, vol. 136, 2017.
- [16] W. Fang, Q. Huang, S. Huang, J. Yang, E. Meng, and Y. Li, "Optimal sizing of utility-scale photovoltaic power generation complementarily operating with hydropower: a case study of the world's largest hydro-photovoltaic plant," *Energ convers manage*, vol. 136, 2017.
- [17] Y. Li, J. Anderson, F. Z. Peng, and D. Liu, "Quasi-Z-source inverter for photovoltaic power generation systems," in *Proceedings of the applied power electronics conference and exposition*, February 2009.
- [18] T. Noguchi, S. Togashi, and R. Nakamoto, "Short-current pulse-based adaptive maximum-power-point tracking for a photovoltaic power generation system," *Electrical Engineering in Japan*, vol. 139, no. 1, pp. 65–72, 2002.
- [19] Y. A. Lawati, J. Kelly, and S. Dan, "Short-term prediction of photovoltaic power generation using Gaussian process regression," 2020.
- [20] X. Ma, Q. Pang, and Q. Xie, "Short-term prediction model of photovoltaic power generation based on rough set-BP neural network," *Journal of Physics: Conference Series*, vol. 1871, no. 1, Article ID 012010, 2021.
- [21] Y. Sun, F. Wang, and Z. Mi, "Short-term prediction model of module temperature for photovoltaic power forecasting based on support vector machine," 2015.
- [22] C. Hang, W. S. Cao, and P. J. Ge, "Forecasting research of long-term solar irradiance and output power for photovoltaic generation system," in *Proceedings of the Fourth International Conference on Computational & Information Sciences*, IEEE Computer Society, Chongqing, China, August 2012.
- [23] H. Cui and L. Yao, "Short term prediction of photovoltaic generation output based on similar days," *Power System Protection and Control*, vol. 43, no. 20, 2016.
- [24] Y. Sun, W. Fei, Z. Zhao et al., "Research on short-term module temperature prediction model based on BP neural network for photovoltaic power forecasting," *Power & Energy Society General Meeting, IEEE*, , July 2015, <https://doi.org/10.1109/PESGM.2015.7286350>.
- [25] P. Lin, Z. Peng, Y. Lai, S. Cheng, Z. Chen, and L. Wu, "Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets," *Energy Conversion and Management*, vol. 177, no. DEC, pp. 704–717, 2018.
- [26] Y. Yoneda and T. Tanaka, "Short term PV prediction using committee kernel adaptive filters," *IEICE technical report. Communication systems*, vol. 111, 2011.
- [27] X. Wang, Y. Sun, D. Luo, and J. Peng, "Comparative study of machine learning approaches for predicting short-term photovoltaic power output based on weather type classification," *Energy*, vol. 240, 2022.
- [28] D. Lee, J. W. Jeong, and G. Choi, "Short term prediction of PV power output generation using hierarchical probabilistic model," *Energies*, vol. 14, 2021.
- [29] P. M. Kumar, R. Saravanakumar, A. Karthick, and V. Mohanavel, "Artificial neural network-based output power prediction of grid-connected semitransparent photovoltaic system," *Environmental Science and Pollution Research*, vol. 29, no. 7, pp. 10173–10182, 2022.
- [30] J. Lin and H. Li, "A hybrid K-Means-GRA-SVR model based on feature selection for day-ahead prediction of photovoltaic power generation," *Journal of Computer and Communications*, vol. 9, no. 11, p. 21, 2021.