Research Article

A Power Forecasting Method for Ultra-Short-Term Photovoltaic Power Generation Using Transformer Model

Fengyuan Tian, Xuexin Fan, Ruitian Wang, Haochen Qin, and Yaxiang Fan

National Key Laboratory of Science and Technology on Vessel Integrated Power System, Naval University of Engineering, Wuhan 430033, China

Correspondence should be addressed to Ruitian Wang; wangrt4321@163.com

Received 5 September 2022; Revised 20 September 2022; Accepted 13 October 2022; Published 28 October 2022

Abstract

The volatility of solar energy, geographic location, and weather factors continues to affect the stability of photovoltaic power generation. Reliable and accurate photovoltaic power prediction methods not only effectively reduce the operating cost of the photovoltaic system but also provide reliable data support for the energy scheduling of the light storage microgrid, improve the stability of the photovoltaic system, and provide important help for the optimization operation of the photovoltaic system. Therefore, it is an important study to find reliable photovoltaic power prediction methods. In recent years, researchers have improved the accuracy of photovoltaic power generation forecasting by using deep learning models. Compared with the traditional neural network, the Transformer model can better learn the relationship between weather features and has good stability and applicability. Therefore, in this paper, the Transformer model is used for predicting ultra-short-term photovoltaic power generation, and the photovoltaic power generation data and weather data in Hebei are selected. In the experiment, the prediction result of the transformer model was compared to the GRU and DNN models to show that the transformer model has better predictive ability and stability. Experimental results demonstrated that the proposed Transformer model outperforms the GRU model and DNN model by a difference of about 0.04 kW and 0.047 kW in the MSE value, and 22.0% and 29.1% of the MAPE error. In addition, the public DC competition dataset is selected for control experiments to demonstrate the general applicability of the transformer model for PV power prediction in different regions.

1. Introduction

Traditional power production consumes fossil fuels such as coal, oil, and natural gas and also leads to environmental pollution in the form of carbon dioxide [1]. As a simple, clean, and safe renewable energy, solar energy has gradually become an important source of electricity generation, which not only has the potential to produce unlimited clean energy but also will certainly bring considerable economic benefits and social benefits. In the past two decades, the popularity of photovoltaic systems in the energy market has continued to increase, and their installed capacity has continued to grow. By July 2021, China's newly installed PV capacity was 17.94 million kW, accounting for 26% of the total newly installed power generation capacity [2]. In addition, China’s photovoltaic power generation industry showed good momentum for development momentum under policy support.

Since photovoltaic power generation mainly depends on solar irradiance, temperature, humidity, and other weather conditions and location conditions, it has strong uncertainty and volatility. The reliable photovoltaic power generation forecast method will not only greatly reduce this uncertainty and enhance the stability of system operation, but also improve the reliability and penetration level of photovoltaic systems, maintain power quality, and improve economic feasibility. Therefore, an accurate photovoltaic power generation power forecast is a hot research topic.

Generally speaking, methods for ultra-short-term photovoltaic power generation power prediction are mainly divided into physical methods and statistical methods [3]:

- Physical method:
  - The physical method relies on physical models established by
2. Related Work

Machine learning methods have been widely used in PV power generation. To improve prediction performance, significant attention has lately been drawn to SVM and deep learning algorithms [8]. For example, Pan et al. optimized the support vector machine (SVM) by using the global search ability of the ant colony algorithm (ACO), which greatly improved the prediction accuracy of the model, but the ant colony algorithm is easy to fall into local optimum [9]. The authors combined the Artificial Bee Colony (ABC) and the Support Vector Machine (SVM) to form the ABC-SVM algorithm. Compared with the traditional SVM algorithm, it has fewer control parameters, stronger optimization ability, and higher prediction accuracy [10].

Li et al. proposed an LSTM-FC deep learning algorithm composed of long-term short-term memory (LSTM) and fully connected (FC) layers, to further study the time correlation for improving the prediction accuracy [11]. The simulation results show that the LSTM-FC is superior to SVM, gradient boosting decision tree (GBDT), generalized regression neural network (GRNN), and feedback neural network (FFNN). However, they only considered the time correlation and did not catch the correlation between the weather data and the PV power generation. Yongsheng et al. established the ELM-LSTM model [12] and used the multi-model univariate extreme learning machine (ELM) to screen out the influential factors with high correlation with photovoltaic power generation, so the hybrid model could better capture the characteristics of information and improve the applicability. This model has the characteristics of fast running speed and low complexity, but it has the defects of easy overfitting and being easy affected by outliers.

Abdel Basset et al. used convolutional layers to redesign the gates of the GRU to enable efficient extraction of position and temporal characteristics in the PV power sequences [13]. Sivakumar et al. used ANN and regression modeling to develop a time series model. In those models, machine learning can help forecast solar energy output [14]. Otherwise, the authors combined LSTM with CNN, wavelet packet decomposition (WPD), wavelet transform (WT), and other methods, and combined the particle swarm algorithm (PSO) with the adaptive neuro-fuzzy inference system (ANFIS) to improve the performance, stability, and reliability of model extraction data features [15–18]. The authors applied the optimal frequency domain decomposition method to deep learning and used correlation to obtain the optimal frequency cutoff points of the decomposition components [19].

As shown in Table 1, the hybrid algorithm proposed by researchers has different advantages and disadvantages. The purpose of the aforementioned methods is as follows: (1) improve the ability of the model to analyze correlations between sample features; (2) improve the stability, reliability, and applicability of the combined model by adjusting the structure. Compared with these combined models, the Transformer model used in this paper has a simpler structure, can meet the experimental requirements, and has good reliability and applicability.
3. Proposed Method

The whole structure of the method in this paper is shown in Figure 1. Firstly, weather feature sequence and PV power data are extracted from the original data, and then the data are preprocessed. Then, the transformer model is utilized for experiments to verify the advantages of the model under different performance indicators. This section mainly introduces the self-attention mechanism and the Transformer model. First, the input features and output of the model are determined. Then the influence of the power generation at the historical moment on the current generation power is considered, and the input vector of the model is preprocessed. Then, the transformer model is utilized for prediction. It improves the performance, stability, and reliability of model extraction data features. The specific calculation process of the self-attention mechanism is as follows:

First, each value of the input vector sequence is mapped to three different spaces, and then the matrix composed of the query vector, key vector, and value vector is obtained as follows: $Q = [q_1, ..., q_N]$, $K = [k_1, ..., k_N]$, and $V = [v_1, ..., v_N]$. Next, as shown in the following equation, each query vector $q_n$ is processed by the key-value pair attention mechanism to obtain the output vector $h_n$.

$$h_n = \sum_{j=1}^{N} \text{softmax}(s(k_j, q_n))v_j,$$

where $n,j \in [1, N]$ is the position of the sequence of output and input vectors, $s(k_j, q_n)$ is the scoring function of attention.

When the scaled dot product is used as the attention scoring function, the output vector sequence is represented as follows:

$$H = V\text{softmax}\left(\frac{K^TQ}{\sqrt{D_k}}\right),$$

where $\text{softmax}(\cdot)$ is a function normalized by column.

Besides, the self-attention model can be extended to the multi-head attention (multi-head self-attention) model to capture different interaction information in multiple different projection spaces, combining the multihead attention model with a feedforward neural network, called the transformer model.

3.1. Self-Attention. Self-attention is an important part of transformer, which is developed from the attention mechanism. The attention mechanism is a mechanism that imitates the human brain to process information and enhances a small part of useful information from a large amount of information to improve the efficiency of the neural network. With the development of deep learning, attention mechanisms have been widely developed and applied in many fields, such as computer vision, natural language processing, and machine translation, which are usually composed of decoder-encoder structures [20]. It can be regarded as a combination function, which strengthens the influence of a key input on the output by calculating the probability distribution of attention [21]. But the attention mechanism ignores the internal features of the task, that is, the relationship between its internal elements [22].

Self-attention sometimes referred to as internal attention, is an attention mechanism that associates different positions of a single sequence to compute a sequence representation. It was first proposed by the Google machine translation team in 2017, which used the query-key-value (QKV) mode to propose an effective modeling method, leading in the field of natural language processing [23]. The basic idea of the self-attention mechanism is to enhance some parts of the input data while reducing others—the motivation is that the network should pay more attention to small but important parts of the data, and general architecture is shown in Figure 2[24].

The specific calculation process of the self-attention mechanism is as follows:

First, each value of the input vector sequence is mapped to three different spaces, and then the matrix composed of the query vector, key vector, and value vector is obtained as follows: $Q = [q_1, ..., q_N]$, $K = [k_1, ..., k_N]$, and $V = [v_1, ..., v_N]$. Next, as shown in the following equation, each query vector $q_n$ is processed by the key-value pair attention mechanism to obtain the output vector $h_n$.

$$h_n = \sum_{j=1}^{N} \text{softmax}(s(k_j, q_n))v_j,$$

where $n,j \in [1, N]$ is the position of the sequence of output and input vectors, $s(k_j, q_n)$ is the scoring function of attention.

When the scaled dot product is used as the attention scoring function, the output vector sequence is represented as follows:

$$H = V\text{softmax}\left(\frac{K^TQ}{\sqrt{D_k}}\right),$$

where $\text{softmax}(\cdot)$ is a function normalized by column.

Besides, the self-attention model can be extended to the multi-head attention (multi-head self-attention) model to capture different interaction information in multiple different projection spaces, combining the multihead attention model with a feedforward neural network, called the transformer model.

### Table 1: Advantages and disadvantages of several PV forecasting hybrid methods.

<table>
<thead>
<tr>
<th>Method</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACO-SVM</td>
<td>It has good accuracy and good global search ability</td>
<td>The ant colony algorithm is easy to fall into the local optimum</td>
</tr>
<tr>
<td>ABC-SVM</td>
<td>It has fewer control parameters, stronger optimization ability, and higher prediction accuracy</td>
<td>There are difficulties with the initialization parameter settings</td>
</tr>
<tr>
<td>LSTM-FC</td>
<td>It further studies the time correlation for improving the prediction accuracy</td>
<td>It only considers the time correlation and did not catch the correlation between the weather data and the PV power generation.</td>
</tr>
<tr>
<td>ELM</td>
<td>This model has the characteristics of fast running speed and low complexity</td>
<td>It has the defects of easy overfitting and easy to be affected by outliers.</td>
</tr>
<tr>
<td>ANFISA</td>
<td>It improves the performance, stability, and reliability of model extraction data features</td>
<td>The algorithm needs to develop new models to improve the prediction accuracy</td>
</tr>
</tbody>
</table>

3.2. PV Power Generation Prediction Based on Transformer. As shown in Figure 3, the entire network architecture of the transformer model consists of a self-attention mechanism and a feedforward neural network (FNN), which are used for self-learning and self-tuning parameters, respectively [25]. In this work, the input is composed of the current weather characteristics and historical power generation data. The core idea is to calculate the relationship between each sample in the input vector and all the other samples, utilize the relationship to reflect the composition of different samples...
to a certain extent, and adjust the weight of each sample through this relationship to obtain more global expressions.

The Transformer model is essentially an encoder-decoder structure. As shown in Figure 3, the encoder on the left contains multiheaded attention; on the right is the decoder, which comprises two multiples of attention. A residual connection and normalization module are also included in each multiattention to prevent network degradation and normalize each layer. During the training stage of the model, the procedure is as follows:

![Figure 1: Method implementation process.](image-url)
**Figure 2:** General architecture for self-attention mechanism.

**Figure 3:** Transformer model structure.
**Step 1.** The input weather characteristics and historical power generation data $x = (x_{t-\tau}, \ldots, x_{t-1}, x_t, y_{t-1})$ are encoded, and then the position information is added to the encoder to append the position to the input sample, which can be expressed in the following equations:

$$PE(p_{ox}, 2i) = \sin \left( \frac{p_{ox} \cdot 2i}{10000^{\frac{2i}{d_{max}}}} \right),$$

$$PE(p_{ox}, 2i + 1) = \cos \left( \frac{p_{ox} \cdot 2i}{10000^{\frac{2i}{d_{max}}}} \right),$$

where $p_{ox}$ is the position of weather features (or historical energy yield) in the input vector, $d$ represents the dimension of input vector $x$, and $i$ represents the dimension of weather features (or historical energy yield) in the input vector.

**Step 2.** Multiattentional mechanism is applied in the encoder, as shown in Equations (5) and (6). Query vector $Q$, key vector $K$, and value vector $V$ are projected by $h$ different learned linear transformations, and the results of multiple different attentional are splintered together to obtain the output of multiattentional. Then is then residually connected and normalized, and then the feedforward neural network layer is calculated, as shown in Equation (7). Finally, the processed samples are input into the decoder.

Multihead $(Q, K, V) = \text{concat}(\text{head}_1, \ldots, \text{head}_h)W^O$, \hspace{1cm} (5)

$$\text{head}_i = \text{Attention}(QW^O_i, KW^K_i, VW^V_i),$$

$$F(X) = \max(0, XW_1 + b_1)W_2 + b_2,$$ \hspace{1cm} (7)

where $W^O$, $W^K$, and $W^V$ are the linear transformation matrix, $b_1$, and $b_2$ are the bias.

**Step 3.** The information matrix in the decoder is processed by multiple attention, feedforward neural network, and normalization to get the output matrix.

**Step 4.** Finally, the relationship between input and output $y_t = f(x_{t-\tau}, \ldots, x_{t-1}, x_t, y_{t-1})$ is obtained by linear transformation and softmax processing.

### 4. The Experiment Design

**4.1. Experiment Process.** The flowchart of the proposed method is shown in Figure 4. First, the meteorological data and photovoltaic power generation data are extracted from the historical data, and the data is preprocessed and divided into the training set and test set. The transformer model is used in the experiment to train and predict the designed four subsets. Comparing methods include the GRU model and the DNN model.

**4.2. Details of the Experimental Data.** Two datasets are chosen for experiments in this work. The first dataset is the real data collected from the upgraded household microgrid in Hebei from March to November 2021 (referred to simply as the “Household microgrid dataset”).

The weather features include hours, north wind speed, surface temperature, surface pressure, total cloud quantity, total sunshine intensity, air temperature, relative humidity, UV intensity, precipitation, snowfall, and dew point temperature. 6600 data points were collected with a one-hour temporal resolution. The household microgrid dataset is divided into three subsets based on different seasons. Subset 1 contains data from March to May (the spring), with a total of 2208 data points. The data from June to August (the summer) is included in subset 2, amounting to 2208 data points. Subset 3 consists of autumn data from September to November, with 2184 data points in aggregate. The second dataset is the public DC competition data set, consisting of weather data and photovoltaic power generation data in 2017 and 2018. The weather features include irradiance, wind speed, wind direction, temperature, humidity, pressure, etc., with a 15-minute temporal resolution. In this dataset, the weather features and the previous photovoltaic generation power are used to input the experiment, and the current photovoltaic generation power is taken as the experiment output.
The details of the dataset and data partition in this paper is shown in Figure 5 and Table 2.

4.3. Data Preprocessing. Due to abnormal instruments for weather measurement, the experimental data has outliers or missing values and needs to be preprocessed. First, outlier handling is performed by setting it to zero when the number of outliers is relatively small. Instead, the outlier is replaced by the average value of the feature when the amount of outlier data is large. After processing outliers and missing values, data with too many large data values in the dataset will affect the results of data analysis. To rescale different features, minimum-maximum normalization is adopted to map feature data to [0,1] as follows:

\[ x^* = \frac{x - \min}{\max - \min} \]  

(8)

4.4. Experimental Settings. The experiment in this paper is implemented on a desktop with Intel Core i5-10400 CPU, 2.90GHz frequency, RTX2060 GPU, and 16 GB RAM. The methods are implemented in Python 3.6.5 with pycharm. According to the size of the datasets, the model parameters are set as shown in Table 3.

In this paper, mean absolute error (MAE) is mainly used as a loss function, as shown in the following equation:

\[ \text{loss} = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i| \]  

(9)

where \( y_i \) is the predicted value, and \( x_i \) is the real value.

5. Case Study

5.1. Evaluation Criteria. Before analyzing the experimental results, to reflect the experimental error directly, the difference between the predicted results and the real value is calculated, and the Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Squared Error (MSE) are selected as the evaluation criteria. Smaller errors indicate higher forecasting accuracy. As shown in equations (10)–(12), MAE is a basic error evaluation index; MSE will square the deviation, which will magnify the error with a large deviation, which could evaluate the stability of a model [26]. MAPE reflects the ratio between errors and real values.

\[ \text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - x_i|, \]  

(10)

\[ \text{MAPE} = \frac{100%}{n} \sum_{i=1}^{n} \frac{|y_i - x_i|}{x_i}, \]  

(11)

\[ \text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2, \]  

(12)

where \( y_i \) and \( x_i \) are the predicted value and the real value, respectively.

5.2. Experimental Results and Analysis. In the first experiment, the household microgrid dataset is first trained and tested to the transformer model, the experimental results are shown in Table 4. To directly reflect the experimental results, 200 consecutive points are randomly selected as shown in Figure 6. It can be seen from the figure that the trend of the predicted values of the three models is consistent with the real values, but the coincidence degree between the predicted values and the real values of the transformer model is better than that of the other two models. Besides, the MAE MSE and MAPE of the Transformer model experimental results are reduced by 43.5%, 19.4%, and 22.0% compared with the GRU model, respectively. 51.1%, 38.9%, and 29.1%, respectively, compared with the DNN model. This indicates that the transformer model achieved the best prediction result among the three methods.

To explain the influence of weather factors on the predicted results, transformer models are trained and tested for the three subsets, respectively. The experimental results are shown in Table 5 and Figures 7–9, where the blue line is the predicted value and the red line is the real value. It is obvious that under the three subsets, the predicted value of the transformer model is basically consistent with the change trend of the real value, and the three evaluation indexes are all within the acceptable range, indicating that the transformer model has good applicability and reliability for different seasons.

As seen from the experimental results, the experimental results of subset 2 are superior to the other subsets under the three evaluation criteria. More specifically, the MAE value of subset 2 is 0.098, which is 0.089 smaller than subset 1 and 0.0441 smaller than subset 3. When under the MSE value, subset 2 is 0.0696 smaller than subset 1 and 0.0281 smaller than subset 3. The MAPE value of dataset 2 is 0.2639 smaller than subset 1 and 0.1749 smaller than subset 3. The MAPE values of the three subsets are too large, as the actual power generation is too small at some time, so the error percentage is too large. The MAPE value of subset 2 is significantly smaller than that of subsets 1 and 3, and the prediction effect of subset 2 is the best when the photovoltaic power is small. The weather features of subsets 1, 2, and 3 are analyzed, as shown in Figure 10. The sunshine intensity and temperature of subset 2 generally remain within a relatively gentle trend. The total sunshine intensity is greater, the temperature is higher and more stable, and it can provide more stable weather conditions for photovoltaic power generation.

Subsequently, the comparison methods such as the GRU and DNN are tested on the three subsets. To demonstrate the superioriety of the proposed method, the experimental results are compared with those of the proposed method, as shown in Table 6. And 200 consecutive data points are randomly selected for plotting in Figure 11. Figure 11 shows that the predicted value of these methods has the same fluctuation trend as the actual value. Among the three models, the prediction results of the transformer model are closest to the actual values. Compared with the GRU model, the MAE and MSE values are reduced by 0.117 and 0.051 at the highest level, and by 0.118 and 0.068 at the highest level on the basis of the DNN model. It is indicated that when a variety of
different weather features are used as input, the transformer model can better learn the correlation between sample features and make more accurate predictions than the GRU model and the DNN model.

As shown in Figure 12, MAE values, MSE values, and MAPE values of subsets 1, 2, and 3 under different models are compared, and the reduction rates of different indicators of the Transformer model compared with the other two models are shown in Table 7. It is obvious that the experimental results of the transformer model are the smallest of all three groups of data in terms of all evaluation criteria. For subsets 2, the prediction accuracy of the transformer model increased dramatically. Compared with the GRU model, the MAE MSE and MAPE values decreased by 54.4%, 55.9%, and 55.9%, respectively. Compared with the DNN model, the MAE MSE and MAPE values decreased by 54.6%, 65.4%, and 41.7%, respectively. Above all, the transformer model has better forecasting ability and more stable performance compared with the other two methods in different seasons.

In addition, to verify the generalization capabilities of the transformer model, the experiment is tested on the DC competition dataset. In the experiment, the parameters of the transformer model are adjusted by the size of the dataset. In detail, the multiattention node number, batch size, number of iterations and the step of sliding value are set as 20, 200, 100, and 1. Similar to the Household microgrid dataset, the GRU and DNN models are chosen for comparison. As depicted in Figure 13, the experimental results of the three models are compared, and the 4000th to 4500th data points are drawn. These results show that the forecasted results of the three models are consistent with the changing trend of the real values. However, it is obvious from Table 8 that the MAE, MSE, and MAPE values of the Transformer model experimental results are 7.2%, 2.5%, and 60.5% lower than those of the GRU model, and 16.4%, 5.9%, and 105.1% lower than those of the DNN model.
lower than those of the DNN model. That means, for the DC competition dataset, the transformer model also achieved better forecasting performance than the GRU and DNN models.

In addition, the experimental results of the DC competition dataset are compared with the household microgrid dataset. It is obvious that under the Household microgrid dataset, the predictive results of the three models are better than those of the DC competition dataset. The reason is that the weather features in the DC competition dataset are fewer than the household microgrid dataset, indicating that reasonable and sufficient data selection and design are necessary before training and prediction.

According to the above experimental results, the transformer model has better forecasting ability than the traditional neural network models.
Figure 8: Experimental results of subset 2.

Figure 9: Experimental results of subset 3.
Figure 10: Sunshine strength and temperature comparison between subsets 1, 2, and 3.

<table>
<thead>
<tr>
<th>Subset</th>
<th>Model</th>
<th>MAE</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subset 1</td>
<td>Transformer</td>
<td>0.187</td>
<td>0.106</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td>GRU</td>
<td>0.256</td>
<td>0.157</td>
<td>1.549</td>
</tr>
<tr>
<td></td>
<td>DNN</td>
<td>0.256</td>
<td>0.153</td>
<td>1.698</td>
</tr>
<tr>
<td>Subset 2</td>
<td>Transformer</td>
<td>0.098</td>
<td>0.036</td>
<td>0.568</td>
</tr>
<tr>
<td></td>
<td>GRU</td>
<td>0.215</td>
<td>0.081</td>
<td>1.289</td>
</tr>
<tr>
<td></td>
<td>DNN</td>
<td>0.216</td>
<td>0.104</td>
<td>0.974</td>
</tr>
<tr>
<td>Subset 3</td>
<td>Transformer</td>
<td>0.133</td>
<td>0.065</td>
<td>0.743</td>
</tr>
<tr>
<td></td>
<td>GRU</td>
<td>0.158</td>
<td>0.079</td>
<td>0.754</td>
</tr>
<tr>
<td></td>
<td>DNN</td>
<td>0.194</td>
<td>0.110</td>
<td>1.032</td>
</tr>
</tbody>
</table>
Figure 11: Experimental results of subsets 1, 2, and 3 in different models (a) subset 1, (b) subset 2, (c) subset 3.

Figure 12: Continued.
Table 7: Different model experimental results error comparison (%).

<table>
<thead>
<tr>
<th></th>
<th>Transformer vs GRU</th>
<th>Transformer vs DNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAE</td>
<td>MSE</td>
</tr>
<tr>
<td>Subset 1</td>
<td>−27.0</td>
<td>−32.7</td>
</tr>
<tr>
<td>Subset 2</td>
<td>−54.4</td>
<td>−55.5</td>
</tr>
<tr>
<td>Subset 3</td>
<td>−15.8</td>
<td>−17.7</td>
</tr>
</tbody>
</table>

(Note. If the value is negative, it means that the evaluation index of the experimental results decreases by the corresponding percentage.)

Table 8: Experiment results of the DC competition dataset.

<table>
<thead>
<tr>
<th>The evaluation index</th>
<th>MAE</th>
<th>MSE</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transformer</td>
<td>1.085</td>
<td>4.042</td>
<td>0.332</td>
</tr>
<tr>
<td>GRU</td>
<td>1.163</td>
<td>4.142</td>
<td>0.533</td>
</tr>
<tr>
<td>DNN</td>
<td>1.263</td>
<td>4.281</td>
<td>0.681</td>
</tr>
</tbody>
</table>
6. Conclusion
Based on the characteristics of photovoltaic power generation input features, a power prediction method for ultra-short-term photovoltaic power generation based on the Transformer model is proposed in this paper. According to the experimental results of the household microgrid dataset and DC competition dataset, it is obvious that the Transformer model has excellent generalization capabilities, which can be well applied to different datasets. Besides, compared with GRU and DNN models, the transformer model can better adapt to the changes in weather characteristics. In addition, through the analysis of the experimental results of three subsets of the household microgrid dataset, the model achieves better prediction results when the sunshine intensity and temperature are relatively stable. The limitations of the proposed methods are that the high computation cost of the training stage leads to huge resource consumption.

In summary, the proposed ultra-short-term photovoltaic power generation forecasting method based on the transformer model has better and more stable time series forecasting ability and generalization ability, which is of great significance for practical engineering applications.

Data Availability
Due to the confidentiality of project data, the household microgrid dataset adopted in this paper cannot be disclosed. The DC competition dataset is available on the web.

Conflicts of Interest
The authors declare that there are no conflicts of interest.

Acknowledgments
This work was supported by the National Natural Science Foundation of China (no. 51907199, 52007196).

References


