County-level Distributed PV Day-ahead Power Prediction based on Grey Correlation Analysis and Transformer-GCAN Model

Pei Zhang1,2, Bin Zhang1, Jinliang Yin3, Jie Shi4

1. College of Electrical Automation and Information Engineering, Tianjin University, Tianjin 300072, China

2.School of Electrical Engineering, Beijing Jiaotong University, Beijing 100091, China

3. School of Electrical Engineering and Automation, Tianjin University of Technology, Tianjin 300384, China

4. School of Physical Science and Technology, University of Jinan, Jinan 250014, China

***Abstract*— The distributed photovoltaic (PV) power stations within the entire county exist spatiotemporal correlation. Merely considering temporal correlation makes it challenging to meet the day-ahead scheduling demands. This paper proposes a distributed PV county-level day-ahead power prediction method based on grey relational analysis and the Transformer-Graph Convolutional Attention Network (Transformer-GCAN) model. Firstly, the grey relational degree is used to measure the relevance between each distributed PV stations, and the connection relationship of the station graph is determined based on the analysis results. Secondly, the Transformer network is utilized to extract the temporal features of each PV sequence in the graph. Based on the Graph Convolutional Network (GCN) model, a Graph Attention Mechanism** **(GAT) is introduced to dynamically extract spatial features between each photovoltaic station in the graph. Finally, the integration of spatiotemporal features is achieved through a fully connected neural network, enabling day-ahead power prediction at the county level. Case analysis results demonstrate that compared with the Transformer-GCN model, the Root Mean Square Error (RMSE) of the power prediction model proposed in this paper is reduced by 11.90%, 15.72% and 19.61% respectively in sunny days, cloudy days and rainy days.**

***Index Terms*—Distributed PV, Transformer, GAT, Day-ahead Power Prediction**

# I. INTRODUCTION

P

hotovoltaic (PV) power generation has experienced rapid development due to its advantages as a clean energy source. PV power stations can be classified into two types: centralized and distributed. Among them, the proportion of installed capacity for distributed PV systems has been increasing year by year [1]. In China, by the end of 2022, the cumulative grid-connected capacity of PV had reached 39,204 MW, with 23,442 MW from centralized PV power stations and 15,762 MW from distributed PV power stations. The proportion of installed capacity for distributed PV has been increasing year by year. However, the dispersed geographical locations of distributed PV stations result in their output characteristics being influenced by their spatial positions. The accuracy of day-ahead prediction is challenging to meet scheduling requirements. Therefore, it is significant task to predict accurately the day-ahead power of distributed PV for the optimization, scheduling, and management of power systems [2].

Currently, centralized PV prediction technology is relatively mature. The PV power prediction methods can be divided into mechanism-driven methods and data-driven methods. The mechanism-driven method utilizes the geographical and meteorological information of PV power stations to establish prediction models based on physical principles[3-4]. Data-driven methods establish prediction models by analyzing the complex mapping relationships between historical power data, meteorological factors, and other data. It mainly includes statistical methods and machine learning methods. Statistical methods primarily include grey theory[5], linear regression[6], time series forecasting[7], and more. These methods have simple models and strong interpretability but with relatively lower prediction accuracy. They are generally suitable for data that follows specific distributions. Machine learning methods primarily include support vector machines[8], extreme learning machines[9], neural networks[10-11], and more. These methods have a higher model complexity and offer greater prediction accuracy compared to statistical prediction methods.

In recent years, deep learning models have been widely studied in the field of PV forecasting[14]. Convolutional Neural Networks (CNN) are typically used for image data. CNN can extract features from images through convolution operations, enabling tasks such as classification or regression[14]. For time-series data such as meteorological data, CNN can process the data by extracting features for PV prediction. Recurrent Neural Networks (RNN) is a type of neural network model used for processing time-series data[15], and PV forecasting is a typical time-series prediction problem. Therefore, RNN have been widely applied in photovoltaic prediction and have achieved good predictive performance. However, when dealing with long time sequences, RNN encounters issues such as vanishing gradients and exploding gradients. To address this issue, Long Short-Term Memory networks (LSTM) was introduced[16]. LSTM addresses the vanishing and exploding gradient problems by introducing gate control mechanisms. In [16], an LSTM model was used to learn the temporal features of the photovoltaic power sequence, while also establishing power prediction models for different weather types. However, when dealing with long time sequences, LSTM exists issues such as lengthy training times and suboptimal convergence in terms of effectiveness. To address these issues, In [13], a CNN model was used to learn the temporal characteristics of shorter segments within the PV time series. However, one-dimensional convolution typically requires adding multiple layers of convolution or enlarging the kernel size to effectively capture the time dependencies in long time series data. So, it may not be suitable for long time series prediction problems. To address this issue, Temporal Convolutional Neural Networks (TCN) was proposed. TCN utilizes dilated causal convolutions to learn temporal dependencies at longer time scales[17]. The application of attention mechanisms in photovoltaic power prediction can effectively improve the model's ability to identify important features, thereby enhancing prediction accuracy. In [18], the Transformer model was used to extract the temporal features of PV sequences, achieving accurate short-term PV power prediction.

The above-mentioned prediction methods were primarily applicable to centralized PV power stations[14-18]. Distributed PV systems are geographically dispersed, and there exists spatial correlation between photovoltaic outputs. Therefore, methods for predicting distributed photovoltaic power need to be improved upon centralized photovoltaic power prediction methods. In recent times, many researchers have adopted cluster-based forecasting methods to overcome the issues of building a large number models[19-21]. Cluster power forecasting mainly includes methods such as the extrapolation method, statistical upscaling method, and accumulation method. When using extrapolation methods for cluster power prediction, historical operational data from each individual station in the cluster is not required. This can lead to data waste, and the final prediction accuracy is influenced by the selection of weights. When using statistical upscaling methods for cluster power prediction, only a few power stations are needed to predict the power for the entire area. However, improper selection of reference stations can directly affect the power prediction results. Accumulation methods require modeling of each distributed photovoltaic user, making them suitable for smaller-scale power stations. The accumulation method not captures the spatial correlations between each PV stations. In [19-20], a cluster-based cumulative prediction method for distributed photovoltaic power was adopted. By dividing the distributed PV cluster, the modeling complexity was significantly reduced, confirming the feasibility of cluster-based cumulative prediction. In [21], a grid partitioning method was applied to large-scale PV regions, predicting the power output for each grid separately. This method effectively enhanced the accuracy and robustness of power prediction for distributed PV station clusters.

However, the above prediction methods rely on the concept of algebraic summation for predicting each photovoltaic station or region. This does not effectively capture the spatial correlation between individual photovoltaic stations. Centralized PV power stations are located in relatively concentrated geographical areas, with distances between each station typically within the range of 5-10 kilometers. The meteorological conditions are the same for each station. However, distributed PV are usually rooftop solar installations, scattered across different geographical locations. Each rooftop PV can be influenced by the surrounding geographic environment, including shading and the movement of clouds. There is spatial correlation in the output characteristics among different rooftop solar installations. Graph convolutional neural networks (GCN) can capture the spatial correlation between different nodes in the picture through the method of graph convolution[22]. For larger geographic areas like counties and towns, it is necessary to use GCN to build models. However, when processing graph data, GCN assign identical weights to different nodes, limiting the ability of the GCN model to extract spatial features. To address this issue, this paper introduces a graph attention mechanism based on the GCN model. The graph attention mechanism assigns different weights to nodes using attention mechanisms. This enables the model to better learn the spatial correlation between different distributed photovoltaic stations.

The paper considers the temporal and spatial correlations of distributed PV power stations, and proposes a distributed PV county-level day-ahead power prediction model based on Transformer-GCAN.

The main innovative contributions of this paper are as follows: 1) introducing a graph attention mechanism based on the GCN model to extract spatial features of different distributed photovoltaic stations; 2) utilizing the self-attention mechanism in the Transformer model to extract temporal features of the nodes in the GCN model.

# II. Materials and Methods

## A. Description of Distributed PV Power Prediction Problem

Distributed photovoltaic power prediction refers to using historical power data collected from distributed photovoltaic stations and meteorological data provided by numerical weather prediction (NWP) to predict the power for the next day. Each distributed PV station forms a graph

 (1)

 (2)

represents the set of each distributed PV station (node), andrepresents the number of stations.represents the set of edges for distributed PV stations, indicating whether there exists a strong spatial correlation between distributed PV stations. represents the adjacency matrix formed by each PV station. The elements in matrixrepresent whether there is an edge or connection between vertexand vertex. The prediction function is as follows.

 (3)

where,represents the distributed PV power prediction function,represents the feature matrix of graphat time ,represents the number of distributed PV stations, andrepresents the number of features for each distributed PV station.

## B. Analysis of Distributed Photovoltaic Correlation Based on Grey Relational Analysis

Grey correlation analysis calculates the correlation coefficient and correlation degree between the comparison series and the reference series, and the calculation equation of the correlation coefficient usually involves the absolute difference between the comparison series and the reference series at each data point, and this difference reflects the degree of deviation between the two at that point. The degree of correlation is weighted by the correlation coefficient of all data points, which reflects the overall degree of correlation between the comparison series and the reference series over the overall series. This correlation value can be used to quantify the degree of similarity and association between variables in spatial locations. To determine the connectivity relationship between distributed photovoltaic stations, the grey relational analysis method is used to select strongly correlated stations for each station. The equation for calculating the correlation coefficient between the *k-*th and *i-*th photovoltaic stations is as follows[5].

 (4)

where,represents the *i-th* power value of *k-th* PV station,represent the length of the sequence of input power for photovoltaic stations, andrepresents resolution coefficient.

Equation (2) is used to calculate the grey relational coefficient between all photovoltaic stations in the county. Based on the calculation results, the photovoltaic stations that are highly correlated with each station are selected to determine the structure of the distributed photovoltaic station graph.

## C. Day-ahead photovoltaic power prediction for distributed photovoltaics based on the Transformer-GCAN model

The framework for distributed photovoltaic power prediction based on the Transformer-GCAN model is shown in Figure 1. The Transformer model is used to extract the temporal features of each photovoltaic sequence. GCAN model is used to extract the spatial features between photovoltaic stations. A fully connected neural network is used to integrate the spatiotemporal features and output the prediction results.

## D. Temporal feature extraction based on the Transformer model

Traditional RNN models encounter vanishing or exploding gradient problems when learning long-term temporal features in time series data. In contrast, the long-term dependency features in time series data are captured by the Transformer model through the utilization of a multi-head attention mechanism for computation.

The multi-head attention mechanism maps the input photovoltaic sequence into queries, keys, and values. By taking the outer product of queries and keys, a weight matrix is obtained. After applying the Softmax activation function, an attention coefficient matrix is derived. Multiplying this matrix with the original input sequence yields an output result influenced by the weights. The model utilizes the multi-head attention mechanism in the Transformer model to learn the temporal attention coefficients of the photovoltaic sequence.

A temporal attention coefficient learning structure based on multi-head attention mechanism is designed, as shown in Figure 2. Temporal attention learning analyzes the impact of information carried by historical moments of the photovoltaic sequence on the prediction results. It assigns different weights to different moments to distinguish the influence of different historical states on the prediction results.



**Fig. 1**. PV Power Prediction Framework based on Transformer-GCAN model



**Fig. 2.** Temporal attention coefficient learning

In Figure 2, represent the information at the *t*-th time step. Three linear transformation matrices , , and  are used to multiply with  to obtain , , and  respectively.represents the influence value of the information at the *n*-th time step on the information at the *t*-th time step. After applying the Softmax function, it becomes the weight value.represents the output value at the *t*-th time step considering the temporal weights. The specific calculation equations are as follows.

 (4)

 (5)

 (6)

 (7)

Where,represents the dimension of vectorand at the time step,represent the length of the time sequence.

## E. Spatial feature extraction based on the GCAN model

CNN input consists of regular grid-like images. As shown in Figure 3, substations with distributed photovoltaic access are geographically dispersed, forming a non-conventional grid graph. Therefore, CNN cannot extract spatial correlation features. GCN are designed to handle graph-structured data. When processing graph data, GCN assign identical weights to different nodes, limiting the ability of the GCN model to extract spatial features. To address this issue, this paper introduces a graph attention mechanism based on the GCN model. Each node in the graph can be assigned different weights based on the features of its neighboring nodes, allowing the model to better learn the spatial correlations between different distributed photovoltaic stations. The principles of GCN and GAT will be introduced below.

(1) Graph Convolutional Network

Distributed rooftop PV with geographically dispersed locations forming an irregular grid. CNN cannot extract spatial correlation features from irregular grid. GCN are designed based on graph theory to perform convolutional operations on topological graphs. According to graph theory, the graph structure is obtained through the Laplacian matrix and eigenvalues. The spectral convolution result on the graph is obtained by convolving the graph signal  with a convolutional kernel function. The convolution process is as shown in Equation (8)[23].

 (8)

where, represents the convolution operator, represents the orthogonal matrix obtained through Laplacian eigenvalue decomposition, and represents the diagonal matrix composed of Laplacian eigenvalues. The Fourier coefficients are obtained by subjecting the graph *G* representing the composition of distributed PV users to a Fourier transform. represents the Laplacian matrix,represents the adjacency matrix of the topological graph *G* composed of distributed PV users, and the diagonal matrix represents the degree matrix. The normalized form of is shown in equation (9).

 (9)

where,  represents the identity matrix.

1. Graph Attention Mechanism

Graph Attention Mechanism is a deep learning model used for processing graph data, particularly suitable for tasks involving learning complex relationships between nodes. Distributed PV stations can be viewed as nodes in a graph. The spatial relationships between these nodes are represented by edges. The GAT network is based on an attention mechanism, allowing the network to dynamically learn the importance of each node in the graph. Unlike traditional GCN, GAT considers the contribution of each neighboring node to the current node, assigning different weights to different neighboring nodes. This allows the model to more effectively capture the complex relationships between nodes. The core of GAT is the utilization of the attention mechanism to learn the weights of relationships between nodes. The calculation equation is as follows.

For each distributed photovoltaic station node *i*, GAT calculates the attention weightbetween it and neighboring distributed photovoltaic station node *j*. The attention weight is used to measure the importance of node *j* to node *i*, and its calculation equation is as follows.

 (10)

Where, *W* represents the weight matrix, *h*i and *h*j represent the feature vectors of nodes *i* and *j*, respectively, and *a* is a learnable attention mechanism function, commonly in the form of a softmax function.

 (11)

Where, *N*i is the set of neighboring nodes of node *i*.

The representation *h*i of each node is updated using the computed attention weights.

 (12)

# III. case study

## A. Case overview

A county in North China contains 10 substations. The number of rooftop PV for each substation is listed in Table 1. The geographical locations of rooftop PV for each substation are shown in Figure 3. The rooftop PV for each substation outputs can be depicted in Figure 4. The PV data for each substation include PV power output, total radiation, total cloud cover, 10-meter temperature, and humidity. The temporal resolution of the PV data is 15 minutes. The time range covers from January 1, 2022, to September 21, 2022.

In this case, the computer configuration is win10 system, 12th Gen Intel @ CoreTM i5-12400, RAM 16GB. The software programming platform is PyCharm, and the prediction model construction work uses the Pytorch framework.

Model input variables are previous day's PV power for each substation, next day's total radiation, total cloud cover, humidity, and temperature data for each power station.

Model output variables are county-level PV power in the next day. In the case, the training set is from January 1 to July 29, 2022, the validation set is from July 30 to August 24, and the test set is from August 25 to September 21, 2022. The prediction time step is 24 hours. The training set is from January 1 to July 29, 2022. The validation set is from July 30 to August 24. The test set is from August 25 to September 21, 2022. The prediction time step is 24 hours. The training set is used to train the model. The validation set is used to select model parameters. The test set is used to test the model performance.

**Table 1.** The number of rooftop PV for each substation.

|  |  |  |  |
| --- | --- | --- | --- |
| Substation | Number | Substation | Number |
| 1 | 106 | 6 | 98 |
| 2 | 136 | 7 | 26 |
| 3 | 14 | 8 | 17 |
| 4 | 86 | 9 | 17 |
| 5 | 145 | 10 | 15 |

To assess the effectiveness of PV day-ahead power forecasting in the county, the experiment employs two metrics: Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

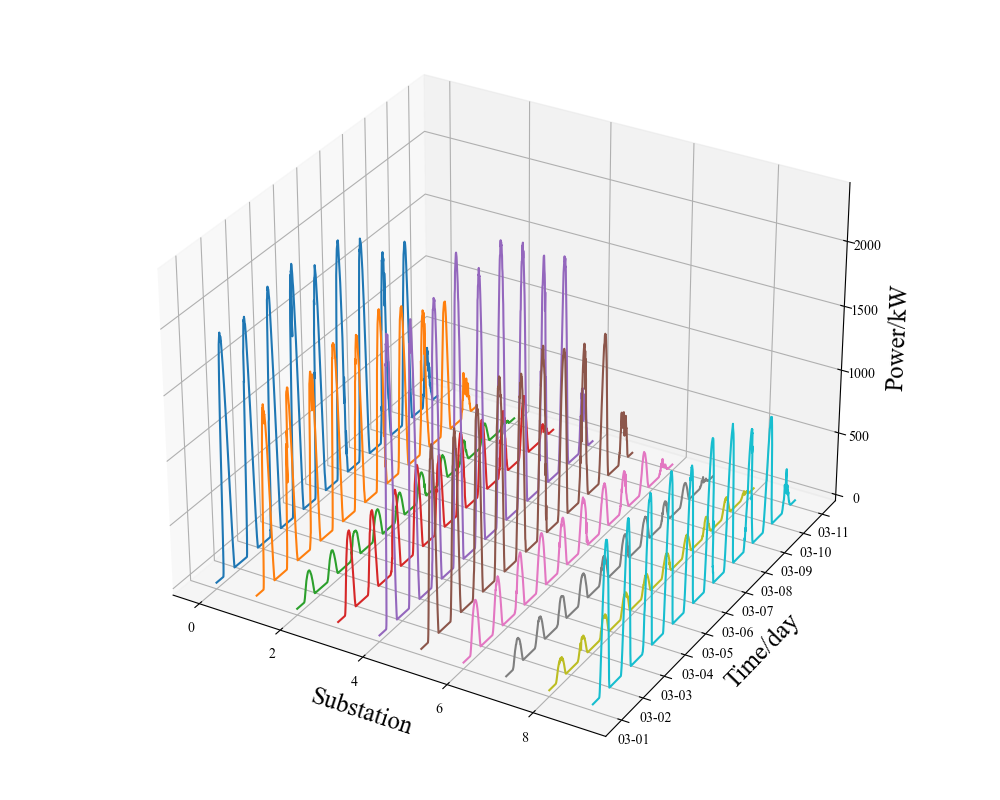
 (13)

 (14)

where,  represents the actual power value,  represents the predicted value.



**Fig.3.** Geographical distribution map of PV.



**Fig.4.** Power curves of each substation outputs (Note: The timeline is a date, for example, 03-01 is March 1.)

The input data for this experiment consists of the power data from the day before the prediction day and the total radiation, total cloud cover, 10-meter temperature, and humidity data for the prediction day. Due to this experiment directly predicting the power of 10 PV stations, the input sequence has a shape of 96×10×5, and the output data is the PV power for the prediction day with a shape of 96×1×1.

## B. Grey correlation analysis results

According to equation (2), calculate the grey correlation between the output power of each distributed photovoltaic station and that of surrounding distributed photovoltaic stations. The calculation results are shown in Table 2 below.

Based on the calculation results in Table 2, when defining the graph of distributed photovoltaic stations, each photovoltaic station is represented as a node in the graph. If the grey correlation between two photovoltaic stations is greater than 0.9, it is considered that there is an edge between the two stations.

## C. Model Parameter Settings

In order to verify the superiority of the Transformer-GCAN model, the Transformer-GCN, TCN-GCN, LSTM-GCN models are used as the experimental control group. The parameter settings and number of training iterations for each model are shown in Table 3-6.



**Fig.5.** Distributed photovoltaic power station undirected graph.

**Table 2.** Grey correlation analysis results

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 1 | 1 | 0.96 | 0.97 | 0.98 | 0.92 | 0.91 | 0.84 | 0.87 | 0.82 | 0.79 |
| 2 | 0.96 | 1 | 0.88 | 0.91 | 0.93 | 0.85 | 0.83 | 0.84 | 0.8 | 0.79 |
| 3 | 0.97 | 0.88 | 1 | 0.91 | 0.88 | 0.94 | 0.87 | 0.85 | 0.83 | 0.80 |
| 4 | 0.98 | 0.91 | 0.91 | 1 | 0.93 | 0.92 | 0.92 | 0.91 | 0.91 | 0.87 |
| 5 | 0.92 | 0.93 | 0.88 | 0.93 | 1 | 0.87 | 0.91 | 0.86 | 0.82 | 0.80 |
| 6 | 0.91 | 0.85 | 0.94 | 0.92 | 0.87 | 1 | 0.95 | 0.96 | 0.93 | 0.91 |
| 7 | 0.89 | 0.83 | 0.87 | 0.92 | 0.91 | 0.95 | 1 | 0.95 | 0.91 | 0.92 |
| 8 | 0.87 | 0.84 | 0.85 | 0.91 | 0.86 | 0.96 | 0.95 | 1 | 0.91 | 0.92 |
| 9 | 0.82 | 0.80 | 0.83 | 0.91 | 0.82 | 0.93 | 0.91 | 0.91 | 1 | 0.92 |

**Table 3.** The parameters of Transformer-GCAN.

|  |  |
| --- | --- |
| Layer | Parameters |
| Input Layer | 5 |
| Transformer Hidden Layer | 32 |
| Dimensions of feedforward neural networks | 64 |
| Number of encoder layers | 2 |
| Number of decoder layers | 1 |
| Transformer output layer | 32 |
| GCN input channel | 32 |
| GCN output channel | 32 |
| GAT input channel | 32 |
| GAT output channel | 32 |
| Dimensions of FC | 32 |
| Output Layer | 1 |
| Number of Training Iterations | 200 |

**Table 4.** The parameters of Transformer-GCN.

|  |  |
| --- | --- |
| Layer | Parameters |
| Input Layer | 5 |
| Transformer Hidden Layer | 32 |
| Dimensions of feedforward neural networks | 64 |
| Number of encoder layers | 2 |
| Number of decoder layers | 1 |
| Transformer output layer | 32 |
| GCN input channel | 32 |
| GCN output channel | 32 |
| Dimensions of FC | 32 |
| Output Layer | 1 |
| Number of Training Iterations | 200 |

**Table 5.** The parameters of TCN-GCN.

|  |  |
| --- | --- |
| Layer | Parameters |
| Input Layer | 5 |
| TCN Hidden Layer | 32 |
| GCN input channel | 32 |
| GCN output channel | 32 |
| Dimensions of FC | 32 |
| Output Layer | 1 |
| Number of Training Iterations | 200 |

**Table 6.** The parameters of LSTM-GCN.

|  |  |
| --- | --- |
| Layer | Parameters |
| Input Layer | 5 |
| LSTM Hidden Layer | 32 |
| GCN input channel | 32 |
| GCN output channel | 32 |
| Dimensions of FC | 32 |
| Output Layer | 1 |
| Number of Training Iterations | 200 |

## D. Comparison and analysis of the prediction results

Establish prediction models based on Tables 3-6, with the predicted results shown in Figure 6.



**Fig.6.** County Prediction Results Comparison Chart.

Figure 6 shows the prediction results of the model for 6 consecutive days. The prediction performance of the model under different weather types is shown in Figure 7-9. The prediction performance metrics for six consecutive days are shown in Table 7.

(1) Comparison and analysis of prediction results under clear sky conditions.

Establish sunny weather prediction models based on Tables 3-6, with the predicted results shown in Figure 7.

According to the analysis based on Figure 7, it can be seen that under clear sky conditions, the fluctuation of the photovoltaic output curve is small. Therefore, the fitting degree of the predicted curves of the four models with the actual curve is good, indicating that the reliability of the four models in predicting the effect under sunny conditions is better.

To more accurately evaluate the prediction performance of the models, two error metrics (RMSE, MAE) are used to characterize the errors of the sunny day predictions. The Transformer-GCAN prediction model is compared with the Transformer-GCN, TCN-GCN and LSTM-GCN prediction models. The comparison results are shown in Table 8.

According to Table 8, it can be observed that under sunny conditions, the Transformer-GCAN model's prediction performance compared to the Transformer-GCN model shows a decrease in RMSE by 11.90% and in MAE by 13.76%; compared to the TCN-GCN model, there is a decrease in RMSE by 17.64% and in MAE by 18.09%; compared to the LSTM-GCN model, there is a decrease in RMSE by 18.61% and in MAE by 18.70%.

1. Comparison and analysis of prediction results under cloudy conditions

Establish partly cloudy weather prediction models according to Tables 3-6, with the predicted results shown in Figure 8.

Based on the analysis in Figure 8, it can be observed that under cloudy weather conditions, the movement of clouds causes differences in cloud cover over each distributed photovoltaic site, thus affecting the solar radiation at ground level for each site. This leads to a decrease in the fitting degree between the predicted curves and the actual curves for the four models. However, the prediction performance of the Transformer-GCAN model is still better than that of the other three models.



**Fig.7.** County Clear-Sky Prediction Results Comparison Chart.

**Table 7.** Prediction performance metrics.

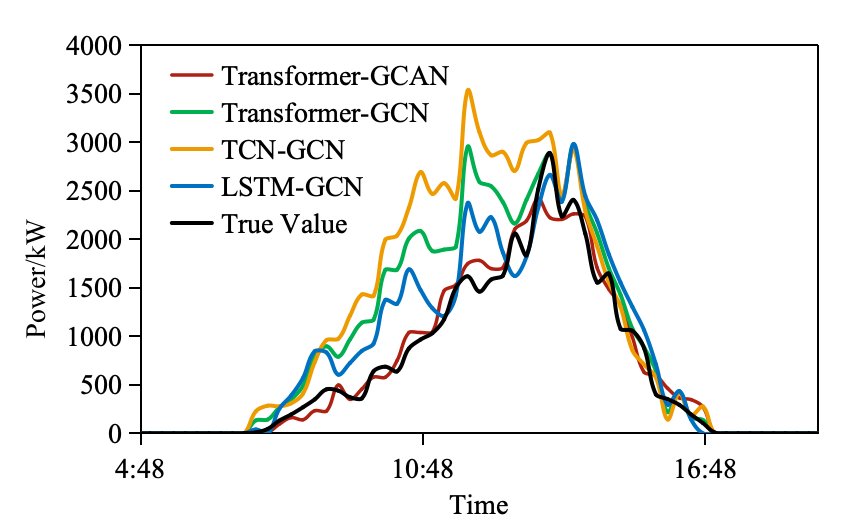
|  |  |  |
| --- | --- | --- |
| **Model** | **RMSE(%)** | **MAE(%)** |
| **Transformer-GCAN** | 6.08 | 3.87 |
| **Transformer-GCN** | 7.41 | 5.31 |
| **TCN-GCN** | 8.82 | 6.01 |
| **LSTM-GCN** | 9.01 | 6.21 |

**Table 8.** Clear sky prediction performance metrics.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **RMSE(%)** | **MAE(%)** | Training Time |
| **Transformer-GCAN** | 4.81 | 3.26 | 163min |
| **Transformer-GCN** | 5.46 | 3.78 | 154min |
| **TCN-GCN** | 5.84 | 3.98 | 137min |
| **LSTM-GCN** | 5.91 | 4.01 | 142min |



**Fig.8.** County Cloudy Sky Prediction Results Comparison Chart



**Fig.9.** County Rainy Weather Prediction Results Comparison Chart

**Table 9.** Cloudy sky prediction performance metrics.

|  |  |  |
| --- | --- | --- |
| **Model** | **RMSE(%)** | **MAE(%)** |
| **Transformer-GCAN** | 6.22 | 3.68 |
| **Transformer-GCN** | 7.38 | 4.99 |
| **TCN-GCN** | 8.31 | 5.46 |
| **LSTM-GCN** | 8.91 | 5.93 |

**Table 10.** Rainy weather prediction performance metrics.

|  |  |  |
| --- | --- | --- |
| **Model** | **RMSE(%)** | **MAE(%)** |
| **Transformer-GCAN** | 8.32 | 5.82 |
| **Transformer-GCN** | 10.35 | 7.51 |
| **TCN-GCN** | 12.27 | 9.27 |
| **LSTM-GCN** | 11.47 | 8.32 |

Further comparison of the four models reveals that the fitting degree of the Transformer-GCAN model's predicted curve with the actual photovoltaic values under cloudy weather conditions is significantly better than that of the Transformer-GCN, TCN-GCN and LSTM-GCN models. This indicates that the Transformer-GCAN model still has high prediction accuracy under non-clear sky conditions. The performance metrics of the four models are shown in Table 9

According to Table 9, it can be observed that under cloudy weather conditions, the Transformer-GCAN model's prediction performance compared to the Transformer-GCN model shows a decrease in RMSE by 15.72% and in MAE by 26.25%; compared to the TCN-GCN model, there is a decrease in RMSE by 25.15% and in MAE by 36.60%; compared to the LSTM-GCN model, there is a decrease in RMSE by 30.19% and in MAE by 37.94%.

(3) Comparison and analysis of prediction results under rainy conditions

Establish rainy weather prediction models based on Tables 3-6, with the predicted results shown in Figure 9.

Based on the analysis in Figure 9, it can be observed that under overcast and rainy weather conditions, rainfall is unevenly distributed throughout the region. This leads to varying degrees of humidity affecting each distributed photovoltaic site, resulting in inconsistent fluctuations in the output of each distributed photovoltaic site. Therefore, the fitting degree between the predicted curves and the actual curves for the four models decreases. However, the prediction performance of the Transformer-GCAN model is still better than that of the other three models.

Further comparative analysis of the four models reveals that the fitting degree of the Transformer-GCAN model's predicted curve with the actual photovoltaic values under overcast and rainy weather conditions is significantly better than that of the Transformer-GCN, TCN-GCN, and LSTM-GCN models. This indicates that the Transformer-GCAN model still has high prediction accuracy under overcast and rainy weather conditions. The performance metrics of the four models are shown in Table 10.

According to Table 10, it can be observed that under overcast and rainy weather conditions, the Transformer-GCAN model's prediction performance compared to the Transformer-GCN model shows a decrease in RMSE by 19.61% and in MAE by 20.50%; compared to the TCN-GCN model, there is a decrease in RMSE by 32.19% and in MAE by 37.22%; compared to the LSTM-GCN model, there is a decrease in RMSE by 27.46% and in MAE by 30.05%.

# IV.Conclusions

This paper proposes a distributed PV day-ahead power prediction method that considers spatiotemporal characteristics. Firstly, it uses grey correlation to measure spatial correlation between distributed PV stations and determines the connectivity of the distributed PV station graph based on the calculation results. Secondly, to address the challenges of modeling and spatial correlation in distributed PV, the method proposes a prediction approach based on the Transformer-GCAN model. This model utilizes the Transformer to extract temporal features of each PV sequence and introduces a graph attention mechanism on top of the GCN model to dynamically extract spatial features between PV stations. Finally, it integrates spatiotemporal features using a fully connected neural network to achieve county-level day-ahead power prediction.

The case analysis results indicate that the Transformer-GCAN model demonstrates higher prediction accuracy under various weather conditions compared to the Transformer-GCN, TCN-GCN, and LSTM-GCN models. Specifically, under sunny conditions, the Transformer-GCAN model's RMSE is reduced by 11.90% compared to the Transformer-GCN model, by 17.64% compared to the TCN-GCN model, and by 18.61% compared to the LSTM-GCN model. Under cloudy conditions, the Transformer-GCAN model's RMSE is reduced by 15.72% compared to the Transformer-GCN model, by 25.15% compared to the TCN-GCN model, and by 30.19% compared to the LSTM-GCN model. Under overcast and rainy conditions, the Transformer-GCAN model's RMSE is reduced by 19.61% compared to the Transformer-GCN model, by 32.19% compared to the TCN-GCN model, and by 27.46% compared to the LSTM-GCN model. It can be inferred that the Transformer-GCAN model is suitable for photovoltaic output prediction under all weather conditions.

High-precision photovoltaic power prediction enables grid operators to more accurately predict changes in photovoltaic power generation trends, thereby optimizing the operational plans of thermal power units in advance. For instance, when a decline in photovoltaic power generation is anticipated, thermal power units can be gradually ramped up to prevent power shortages caused by insufficient ramp rates. Conversely, when an increase in photovoltaic power generation is predicted, the output of thermal power units can be reduced in advance to avoid overgeneration. This coordinated optimization not only reduces the frequent start-stop cycles and ramp pressure on thermal power units, extending their operational lifespan, but also significantly lowers fuel consumption and carbon emissions, while enhancing grid stability and economic efficiency. Furthermore, accurate prediction facilitates better integration of photovoltaic power into the grid, reduces curtailment, and improves the utilization rate of renewable energy, thereby accelerating the transition towards a greener energy structure.

References

[1]Zhang J.J.; Hong L.Q.; Lbrahim. Short-term prediction of behind-the-meter PV power based on attention-LSTM and transfer learning. IET Renewable Power Generation. 2023.

[2]Zhang TF.; Ma FM.; Peng C. A Very-Short-Term Online PV Power Prediction Model Based on RAN with Secondary Dynamic Adjustment. IEEE Transactions on Artificial Intelligence. 2022, 8, 1-1.

[3]Gandoman F H, Abdel Aleem S H E, Omar N, et al. Short-term solar power forecasting considering cloud coverage and ambient temperature variation effects[J]. Renewable Energy, 2018, 123: 793-805.

[4]Saint-Drenan Y M, Good G H, Braun M. A probabilistic approach to the estimation of regional photovoltaic power production[J]. Solar Energy, 2017, 147: 257-276.

[5]Lu H.J.; Chang G.W. Enhanced Grey Theory System Model with Weight Table for Now-casting PV Power Generation Forecast. International Conference on Innovative Smart Grid Technologies. 2018, 18, 952-956.

[6]Kaushik, A.R.; Padmavathi, S.; Gurucharan, K.S.; Raja, S.Charles. Performance Analysis of Regression Models in Solar PV Forecasting. 2023 3rd International Conference on Artificial Intelligence and Signal Processing. 2023.

[7]Leva S, Dolara A, Grimaccia F, Mussetta M, Ogliari E. Analysis and validation of 24 hours ahead neural network forecasting of PV output power. Math Comput Simul. 2017, 131, 88-100.

[8]Mohammadi K, Shamshirband S, Tong CW, Arif M, Petković D, Ch S. A new hybrid support vector machine–wavelet transform approach for estimation of horizontal global solar radiation. Energy Convers Manage. 2015, 92, 162–71.

[9]Nayak, Niranjan; Pani, Alok Kumar. A short term forecasting of PV power generation using couple based particle swarm optimization pruned extreme learning machine. International Journal of Renewable Energy Research. 2019, 9, 1190-1202.

[10]Zhu H, Li X, Sun Q, Nie L, Yao J, Zhao G. A power prediction method for PV power plant based on wavelet decomposition and artificial neural networks. Energies. 2015, 9-11.

[11]De Barros Silva A.W.; Freitas B.B. Methodology Based on Artificial Neural Networks for Hourly Forecasting of PV Plants Generation. IEEE Latin America Transactions. 2022, 20, 659-668.

[12]V. Venugopal; Y. Sun; A. R. Brandt. Short-term solar PV forecasting using computer vision: The search for optimal CNN architectures for incorporating sky images and PV generation history. J. Renew. Sustain.Energy. 2019, 11, 6.

[13]Tang D.Q.; Zhu W.; Hou L.C. Ultra-short-term PV power prediction based on CNN-LSTM-XGBoost model. Chinese Journal of Power Sources. 2022, 46, 1048-1052.

[14]Ren, X.; Zhang, F.; Yan, J.; Liu, Y. A Novel Convolutional Neural Net Architecture Based on Incorporating Meteorological Variable Inputs into Ultra-Short-Term Photovoltaic Power Forecasting. Sustainability 2024, 16, 2786.

[15]Akhter M.N.; Mekhilef S.; Mokhlis H. An Hour-Ahead PV Power Forecasting Method Based on an RNN-LSTM Model for Three Different PV Plants. 2022, 15, 6.

[16]Li F.J.; Wang L.; Zhao J. Distributed PV short-term power prediction method based on weather fusion and LSTM network. Electric Power. 2022, 55, 149-154.

[17]Lin, X.; Niu, Y.; Yan, Z.; Zou, L.; Tang, P.; Song, J. Hybrid Photovoltaic Output Forecasting Model with Temporal Convolutional Network Using Maximal Information Coefficient and White Shark Optimizer. Sustainability 2024, 16, 6102..

[18]López Santos M, García-Santiago X, Echevarría Camarero F, et al. Application of temporal fusion transformer for day-ahead PV power forecasting[J]. Energies, 2022, 15(14): 5232.

[19]Lu J.J; Cai T.; Lang J.X. Short-term Power Output Forecasting of Clustered PV Solar Plants Based on Cluster Partition. Gaodianya Jishu/High Voltage Engineering. 2022, 48, 1943-1951.

[20]Wang X.X; Yu M.; Huo Z.J. Short-term power prediction of distributed PV power plant cluster based on neighbor propagation clustering and LSTNet. Automation of Electric Power Systems. 2023, 47, 133-141.

[21]Liu X.Y.; Wang J.; Yao T.C. Ultra-short-term distributed PV power prediction based on satellite remote sensing. Transactions of China Electrotechnical Society, 2022, 37, 1800-1809.

[22]Zhang M.; Zheng Z.; Liu N. Optimal Graph Structure Based Short-Term Solar PV Power Forecasting Method Considering Surrounding Spatio-Temporal Correlations. IEEE Transactions on Industry Applications. 2023, 59, 345-357.

[23]DAUPHIN Y N; FAN A; AULI M. Language modeling  with  gated  convolutional  networks. 34th International  Conference  on  Machine  Learning. Sydney:IMLS. 2017, 1551-1559.