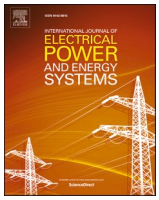
Electrical Power and Energy Systems 172 (2025) 111084



Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes





Optimizing regional PV power forecasting using satellite cloud images and transfer learning strategy

Yang Xie [a,b], Jianyong Zheng [b,*], Mei Fei [c], Gareth Taylor [d]

[a] School of Cyber Security, Guangdong Polytechnic Normal University, Guangzhou 510665, China
[b] School of Cyber Science and Engineering, Southeast University, Nanjing 210096, China
[c] College of Electrical and Power Engineering, Hohai University, Nanjing 210010, China
[d] Brunel Interdisciplinary Power System Research Centre, Brunel University London, London UB8 3PH, UK

ARTICLE INFO

Keywords:
Regional PV power forecasting
Satellite cloud imagery
Attention mechanism
Texture features
Transfer learning

ABSTRACT

As renewable energy, particularly regional photovoltaic (PV), becomes more prevalent in the power grid, accurate forecasting of its output is paramount for its efficient operation. However, challenges persist, including the lack of reliable data, inappropriate data usage, and the computational burdens stemming from the vast number and dispersed nature of regional PV installations. To address these problems, a regional PV power forecasting based on transfer learning and satellite cloud imagery is proposed. Firstly, an algorithmic architecture composed of gray-level co-occurrence matrix and random forest is established for extracting texture features (TF) from satellite cloud images and selecting the TF with the highest correlation to irradiance. Furthermore, an attention mechanism based on long short-term memory is employed to reconstruct these significant TF. These reconstructed TF are then integrated into the training data for the forecasting model, aiming to enhance the correlation between the TF and the forecasting outcome. Finally, a transfer leaning structure combine convolutional neural network and informer is taken as the forecasting model, and the self-organizing map and maximum mean discrepancy algorithm is utilized to distinguish the source and target PV stations. Both the single PV located in UK and the regional PV station located in China are analysis to verify the effectiveness, and several benchmark forecasting methods have been compared, the forecasting approach in this research demonstrated superior performance. With this strategy, the maximum improvement in accuracy reached to 8.7 in term of RMSE, the volume of training data is reduced to 60%, and the computational time reduced to 54.4%.

1. Introduction

1.1. Research background

With the continuous advancement of global renewable energy technologies, the installed capacities of new energy sources, primarily PV power, are on a constant rise [1]. In 2024, global photovoltaic (PV) installation capacities are expected to reach 491 GW, representing a year-on-year increase of 15 % from 2019. This trend is projected to continue, with installations anticipated to reach 565 GW by 2025 [2,3]. Moreover, the growth rate of installed capacity for regional PV is very rapid, resulting in a new energy development strategy where centralized and regional PV grow side by side [4]. The accurate forecasting of the PV power is essential to improve the capacity of the PV in the electricity power system [5]. However, the forecasting results are affected by many factors, such as the weather condition, the movement and types of clouds above the PV power stations [6]. Additionally, due to the surge in the number of PV stations within a region, forecasting each station individually would result in substantial computation time. Such an approach is not only inefficient but also demands significant computational resources [7]. In order to harness the full potential of renewable resources and augment the share of PV power within the grid, it becomes imperative to implement strategies that enhance both forecasting accuracy and computational efficiency [8].

1.2. Literature review

To bolster forecast accuracy, optimizing meteorological data is paramount. Notably, the dynamics and types of clouds overlying PV stations substantially impact their output. This is attributed to clouds

* Corresponding author.
E-mail address: jiany_z135@163.com (J. Zheng).

https://doi.org/10.1016/j.ijepes.2025.111084
Received 4 December 2024; Received in revised form 5 May 2025; Accepted 4 September 2025
Available online 3 November 2025
0142-0615/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (http://creativecommons.org/licenses/by-nc/4.0/).

attenuating the irradiance reaching the PV panels [9]. The movement and types of cloud are captured by local-sensing and remote-sensing cloud monitoring equipment [6]. For the local-sensing devices, the sky imager is employed, and the cloud data over the PV station is fully obtained with the help of the fisheye lens [10]. Then, a series of image of processing is performed on the ground cloud map to predict the future short-term PV power. In order to make the PV power forecasting more accurate, a series work have been done to enhance the ability of capturing the dynamics and types of the cloud based on the equipment of sky imager [11]. For instance, Jiang, J. et al. [12] explore ground-based remote sensing with high temporal and spatial resolution for PV power forecasting, particularly under cloudy conditions. They use sky images captured every minute to determine the distribution of clouds with varying optical depths, and to generate a two-dimensional cloud map. Hu, K., et al. [13] propose a ground-breaking method for cloud motion forecasting using ground-based total sky imager data, enhancing solar PV panel output prediction through advanced digital image processing and the development of an ultra-short-term PV power forecasting model. While methods employing local-sensing devices have been instrumental in capturing and analyzing clouds to enhance PV power forecasting, the acquisition of ground-based cloud data necessitates hardware infrastructure [14], rendering the approach economically challenging.

Addressing this limitation, several studies have turned to public satellite cloud imagery as an alternative to sky cloud images, adopting what is commonly referred to as the remote-sensing technology [15]. For example, Wang, F. and Li, J. [6] introduced a method that leverages the spatiotemporal correlation between multiple PV plants, emphasizing the significance of cloud information from neighboring plants. This approach extracts global cloud distribution information from satellite images to refine forecasting accuracy. In a complementary vein, another study by Wang et al. [16] combined fluctuation pattern recognition with deep learning, extracting cloud features using convolutional autoencoders. These pioneering methodologies underscore the potential of satellite imagery in revolutionizing PV power forecasting. However, while the additional features were extracted directly from the satellite images, which provide an effective method to enhance forecasting performance—especially under changing cloud cover conditions at individual PV sites—the spatial and temporal correlations in satellite cloud images are challenging to capture adequately for regional PV sites using these approaches. To address this challenge, a novel satellite image encoder was designed in [17], utilizing a dual-branch spatiotemporal vision transformer to compress large-scale cloud features into those relevant to the region of interest. The satellite image features are then combined with PV time-series features using a cross transformer with rotary position embedding. The proposed method was validated using data from ten PV stations, demonstrating significant improvement in forecasting performance. However, these approaches neglect the analysis of texture features in satellite cloud imagery and their varying correlations with irradiance, failing to clearly convey the intricate relationship between satellite cloud imagery and irradiance levels. While the forecasting performance for individual PV power systems can be significantly improved, the effect is less pronounced for regional PV power forecasting. This is due to the fact that the correlations of weather conditions among the regional PV sites are not fully captured by the original satellite images, leading to inadequate provision of dynamic weather information across these sites.

Recently, the gray level co-occurrence matrix (GLCM) method has been employed to extract texture features and combined with various machine learning algorithms to capture correlations among multiple PV sites, thereby enhancing the forecasting performance of regional or distributed PV power. For instance, the research [18] proposes the pattern classification and particle swarm optimization optimal weights-based method for calculating cloud motion speed in solar PV power forecasting, referred to as PCPOW. The k-means clustering method, along with TF derived from the GLCM, is employed to classify the clouds. A corresponding combined calculation model is then developed to determine cloud motion speed. The results demonstrate that both the cloud classification and the optimal combination model are effective, and that PCPOW significantly enhances the accuracy of displacement calculations. Furthermore, the research [19], a detailed texture feature map for significant areas was generated using GLCM and local binary patterns (LBPs). This map serves as input for an enhanced ConvNeXt classification network, which accurately categorizes clouds based on meteorological standards. Subsequently, a support vector machine (SVM)-Informer model was constructed to predict the power output of high-efficiency PV systems for the next 5 min. Although these methods show significant improvement in regional PV power forecasting accuracy, the forecasting horizon remains too short to meet the demands of today's power systems, particularly given the increasing proportion of PV power. To address this limitation, we develop a novel PV power forecasting method.

In terms of enhancing forecasting efficiency for regional PV power stations, the primary approach involves the design of an apt forecasting model. In recent years, models grounded in machine learning and deep learning have emerged as the preferred solutions to this challenge. However, the diverse location and various meteorological information have brought challenges to real-time distributed forecasting. Therefore, it is more difficult for forecasting regional power than centralized power. To address this challenge, research [20] propose a four-stage space–time hybrid method for power forecasting of regional PV plants. Firstly, Global and Local Feature Fusion Network (GLFFNet) is proposed to forecast the power output of centralized PV plants. Secondly, the weather clustering type is obtained to get ready for studying the relationship between centralized and regional PV plants. Then, uncovered coefficients are figured out to show the distribution features of PV power both in the time and space domain. Finally, the correlation models are established based on Copula analysis. To reduces computational complexity, research [21] propose PerfCNN-LSTM, a novel model that integrates a CNN and a LSTM network with the Performer self-attention mechanism to enhance PV power generation forecasting. It extracts local features while capturing global features through the Performer layer, which transforms nonlinear attention weight calculations into linear ones. However, the current research ignores how to reasonably divide the sub-regions with similar output characteristics and mine the spatial and temporal correlation between different sub-regions. Therefore, this paper [22] proposes a short-term regional distributed PV power forecasting method based on sub-region division considering spatio-temporal correlation. Firstly, the representative power plant is selected after dividing the sub-region by the AP clustering algorithm. Then, the GCN is used to extract spatial correlation features, and the LSTM is used to extract the evolution features of dynamic spatial correlation features, and the power forecasting models of representative plants in different weather types are established. Although these methods have performed well in regional PV power forecasting, several research gaps remain, including implementation complexity, high computational resource demands, and a need for advanced data processing and feature engineering expertise. These factors can hinder their practical deployment in scientific research [23] Fig. 1.

1.3. Motivation and contributions

Accurate and efficient forecasting of regional PV power is essential for guiding power grid dispatch and enhancing the integration of renewable energy into the new power systems. Thus, research into regional PV power forecasting is imperative. Much recent research has incorporated data from satellite cloud imagery into the training datasets of forecasting models to enhance forecasting accuracy. Different TF of the satellite cloud imagery has variable correlation with the irradiance, so taking the important TF as part of the training data of the forecasting model can significantly improve the forecasting accuracy. Besides, the number of the PV station in special region is huge and it will take long

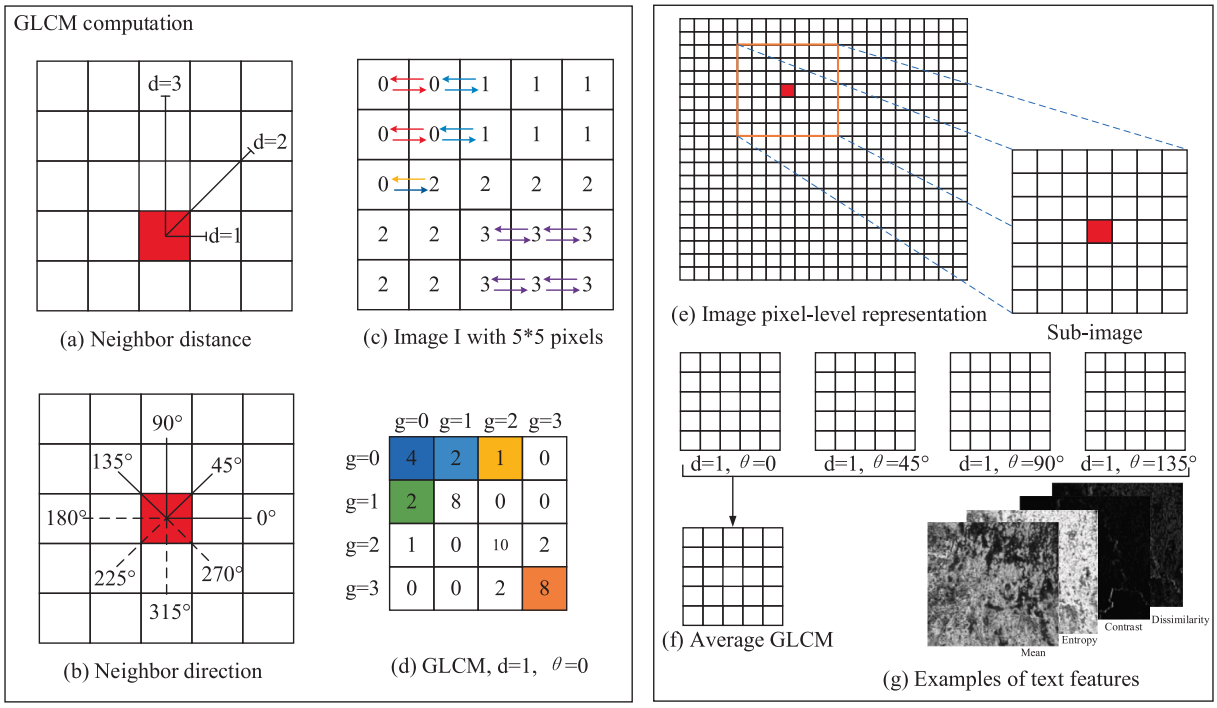


Fig. 1. The computational procedure for the various TF.

computation time to forecast all the PV stations if do not taking consideration of the relationship of different PV stations and use the traditional forecasting model in the region [24]. However, most of forecasting method take the distribution feature of the cloud [6] and extract the relationship between different PV station with the cloud [16], which can not exactly describe the correlation between the cloud and the irradiance and also will take a lot computation time to forecast regional PV power stations. Furthermore, the texture features extracted from satellite images are directly utilized as feature data, which can significantly enhance forecasting accuracy but often overlook the forecasting horizon. Additionally, in the field of regional PV power forecasting, traditional hybrid deep learning methods require training and updating all parameters. Although this process is comprehensive, it leads to significant computational demands and reduced efficiency, posing a major challenge in practical applications.

To fill this gap, a regional PV stations forecasting method based on the TL and satellite cloud imagery is proposed in this research. Firstly, a TF extracting algorithm combined of GLCM and random forest (RF) is utilized to calculate the important TF and analysis the correlation between different TF and irradiance. Additionally, an LSTM with AM is employed to reconstruct the important TF and take the reconstructed TF as part of the training data of the forecasting model. Then, the forecasting model based on TL combined CNN and Informer (INF) is training by the source PV station dataset and relevant model parameters are locked and fine-tuned to meet the forecasting requirements of target PV power stations. During this process, the self-organizing maps (SOM) and maximum mean discrepancy (MMD) algorithm is initially utilized to cluster and analyze the relevance magnitude of all regional PV power stations covered by the cloud layer within the vicinity. Following this, data partitioning is established within this vicinity, using PV power stations with highest correlation and sufficient data as the source domain and the other power stations as the target domain.

To sum up, the forecasting method consisted of three parts. The first is to extract and select the important TF of the satellite cloud imagery. And the second is reconstruct these important TF and take it as part of the training data of forecasting model. Finally, a regional PV power forecasting model based on TL and combined with CNN and INF is proposed. And a series of data in different area and different weather condition are trained by these model and benchmark model to verify the efficiency of the method proposed in this research.

The main contributions of this paper can be summarized as:

1. This paper proposes an architecture that integrates GLCM, RF, and LSTM-based AM algorithms to extract, select, and reconstruct TF from satellite cloud images. These additional data enhance the forecasting performance by improving the correlation between the training data and the output of the PV power station. Notably, this approach significantly boosts the accuracy of PV power forecasting, particularly under varying cloud weather conditions.
2. A transfer learning strategy based on a CNN-INF hybrid deep learning model is proposed, effectively capturing the spatiotemporal correlations within the feature datasets and accelerating the model training process. This method enhances forecasting accuracy across various weather conditions and significantly improves computational efficiency, especially when the number of PV stations in a specific region is large.
3. The effectiveness of the proposed approach is validated using real-world datasets from both individual PV and regional PV stations. When compared to multiple benchmark forecasting methods, our research method exhibits superior performance.

In the remining section of this paper is organized as fellow. Section 2 primarily discusses the proposed methodology of the regional PV power forecasting. In section 3, a series of experiments based on real-data are carried out. Conclusion are drawn in section 4.

2. Methodology

This section will elucidate the methodologies and theoretical underpinnings employed in the advanced regional PV power forecasting framework. Initially, the process of TF extraction from satellite cloud imagery and the subsequent reconstruction of pivotal TF will be delineated. These reconstructed TF furnish supplementary data for the regional power forecasting model, significantly enhancing its accuracy

by incorporating considerations of cloud motion's impacts. Subsequently, a novel hybrid forecasting model that synergizes CNN and INF through a transfer learning strategy is introduced. Within this model, the CNN is tasked with capturing complex feature concatenations, while the INF, structured around multi-head attention mechanisms, augments feature processing capabilities. Moreover, the application of transfer learning strategy substantially enhances computational efficiency across regional PV stations.

2.1. TF extraction and selection from satellite cloud imagery

This phase is dedicated to developing a significant data feature that enhances the precision of PV power forecasting. The influence of clouds on irradiance is predominantly manifested through the clouds' thickness and their diverse TF [25,26]. To extract these TF from satellite cloud imagery, the GLCM and its associated computational methodology are employed [27]. Furthermore, the RF algorithm is introduced in this section to identify highly relevant TF, thereby mitigating feature redundancy. This approach ensures that only the most impactful TF are considered, streamlining the forecasting model for improved accuracy and efficiency.

2.1.1. GLCM

The GLCM was proposed by R. Haralick and others in the 1970 s [28]. It was introduced under the assumption that the spatial distribution relationship between pixels in an image contains texture information, presenting a widely applicable method for texture analysis [29]. The GLCM quantifies the frequency at which particular combinations of pixel intensity values appear in close proximity within an image. In essence, it encapsulates various textural attributes, including contrast, correlation, energy, and homogeneity. The matrix offers a comprehensive depiction of the image's texture, with numerous metrics derived from it serving as representatives of distinct textural features.

Its definition is that starting from pixel with a gray level of i, the probability that another pixel at a distance of d_x, d_y has a gray level of j. In simple term, it can be described as the number of times a certain shape of pixel pairs appears in the entire grayscale image. The mathematical expression can be described as [30]:

$$P(i,j|d,\theta) = \#\{(x,y)\,|f(x,y)=i, f(x+d_x,y+d_y)=j; x,y = 0,1,2,\cdots,N-1\} \quad (1)$$

where, d is the row-pair distance represented by the number of pixels, and typically considers four directions, namely 0°, 45°, 90°, 135°. # represents the set, (x,y) are the pixel coordinates in the image, L is the number of gray levels in the image, and $i,j = 1,2,\cdots,L-1$. Fig. 2 (a)-(d) presents a detailed diagram of the computation of the GLCM. Fig. 2(a) illustrates the varying neighbor distances (d), while Fig. 2(b) depicts commonly utilized directions (θ). Fig. 2(c) presents a specific image of 5×5 pixels at a gray level of 4. Fig. 2(d) showcases the completed GLCM, where the entry $P(0, 0) = 4$ indicates the count of neighboring pixel pairs with gray levels of (0, 0), as demonstrated by blue arrows. This is under the conditions of a neighbor distance (d) of 1 and an orientation (θ) of 0 degrees.

2.1.2. Calculation of TF

The GLCM offers insights into the direction, distance, magnitude, and rate of grayscale transitions within an image. Nonetheless, it does not inherently furnish distinct texture attributes. Consequently, the extraction of statistical measures from the GLCM becomes essential to provide a quantitative representation of TF.

Haralick proposed 14 distinct features from the GLCM, widely utilized for texture analysis in images. These features encapsulate various texture aspects like contrast, uniformity, and homogeneity. The 14 Haralick features include: energy, contrast, correlation, variance, homogeneity, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, maximal correlation coefficient, cluster shade, and cluster prominence. In remote sensing image analysis, commonly employed statistical measures for texture extraction include mean, variance, homogeneity, contrast, dissimilarity, entropy, energy, and correlation. Their expressions are as follows:

$$Mean = \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(i,j,d,\theta) \quad (2)$$

$$Var = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} (i-\mu)^2 f(i,j,d,\theta) \quad (3)$$

$$Hom = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \frac{f(i,j,d,\theta)}{1+(i-j)^2} \quad (4)$$

$$Con = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} (i-j)^2 f(i,j,d,\theta) \quad (5)$$

$$Dis = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} |i-j| f(i,j,d,\theta) \quad (6)$$

$$Ent = -\sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(i,j) \lg f(i,j,d,\theta) \quad (7)$$

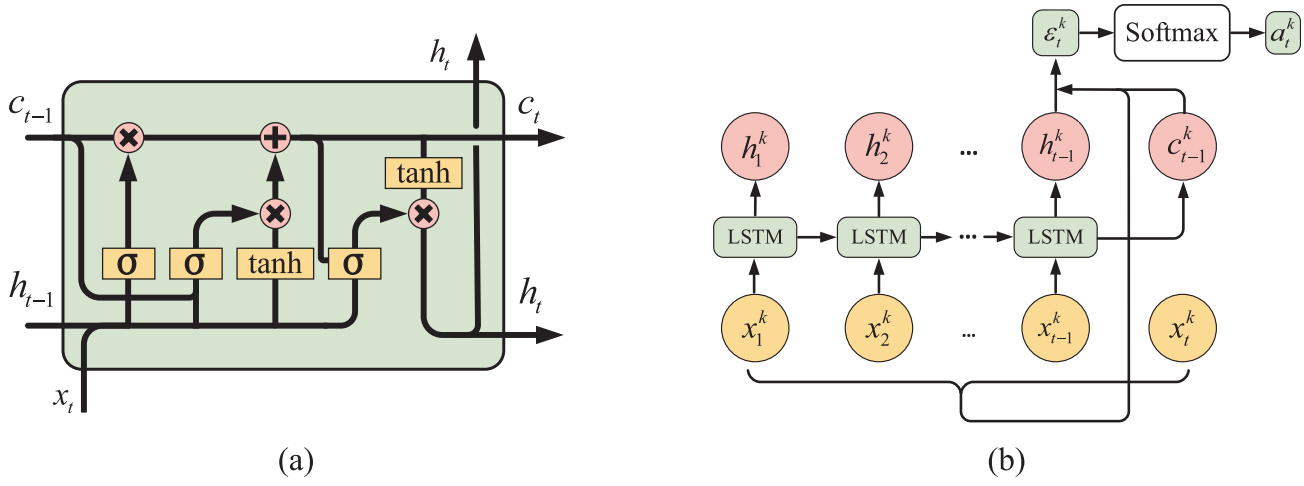


Fig. 2. Architectural schematic of the LSTM-based AM.(a) LSTM unit. (b) The structure of attention layer.

$$Ene = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} f(i,j,d,\theta)^2 \tag{8}$$

$$Cor = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} \frac{(i-\mu)(j-\mu)f(i,j,d,\theta)^2}{\delta^2} \tag{9}$$

$$AutoCor = \sum_{i=0}^{M-1}\sum_{j=0}^{N-1} (ij)f(i,j,d,\theta) \tag{10}$$

where, M and N represent the size of the GLCM, typically $M = N = L$. The computational procedure for extracting various TF following the GLCM computation is detailed in Fig. 1(e)-(g). Fig. 1(e) displays a sub-image representation extracted from the cloud image, highlighting specific pixel characteristics. Fig. 1(f) illustrates the process of computing the average GLCM, emphasizing the methodological steps involved. Finally, Fig. 1(g) presents an example of a texture feature images, which is derived using specific correlation formulas.

2.1.3. Selection of TF

Remote sensing satellite cloud images contain a substantial number of TF. Utilizing all TF to represent the influence on irradiance can not only increase the computational burden of the model but also introduce feature redundancy. This redundancy can, in turn, diminish forecasting accuracy and have other negative effects. To isolate the principal TF from the plethora of TF, feature screening is essential. Given the RF algorithm's prowess in handling high-dimensional data and its capability to evaluate the importance of individual features, coupled with its robust resistance to interference [31], this study opted to use the RF algorithm for screening the TF of satellite cloud images.

When analyzing the impact of different TF in satellite cloud images on irradiance, the RF method can be used for variable importance ranking. The variable importance ranking in RF is based on the predictive power of each feature in the model. In random forest models, variable importance is typically assessed using mean decrease in mean squared error (IncMSE) for regression, and increase in node purity (IncNodePurity) for classification [32,33]. Given the focus of this study on power forecasting for PV stations, IncMSE is the appropriate metric for selecting significant transfer functions. The IncMSE can be expressed as:

$$\text{IncMSE} = \frac{1}{n}\sum \left(y(i) - y_p(i)\right)^2 \tag{11}$$

where, n is the number of samples in the node, y_i is the target value of the i^{th} sample in the node, and $\hat{y}$ is the average target value of all samples in the node. To further demonstrate the distribution of importance for different TF and the selection of significant TF, partial dependence plots (PDP) are utilized here as a method for visualizing variable importance. PDP is a technique for visualizing and interpreting machine learning models. It illustrates the dependency between the target variable and one or multiple feature variables. In essence, a PDP plots the effect of a single feature on the predicted outcome of a model, keeping all other features constant [34]. It captures the relationship between the specified feature and the target variable, allowing us to visualize how the average forecasting changes when a feature varies over its range.

2.2. TF reconstruction via AM

The contribution of different TF to irradiance measurements varies significantly. To address this, an attention mechanism is employed to selectively reconstruct these TF. The AM is a computational model that mimics human visual and cognitive attention processes. The AM allows models to assign different weights and attention levels to different parts of the input when processing sequences, thus concentrating more on information relevant to the task. Within the TF of cloud, the influence on irradiance varies among the features. Certain TF exert a pronounced impact on irradiance, whereas others present a more subdued effect. Therefore, by using the AM to reconstruct different TF, it can help the model better capture the relationships and importance between input features, enhancing the performance of the model. In this research, a reconstruction of different TF is carried out based on LSTM as the basic structure.

LSTM, a type of recurrent neural network, is specifically designed for modeling sequence data. It can capture the temporal correlations and long-distance dependencies in the input sequence. When combined with the AM, LSTM can better capture the relationships between variables, and the AM allows the model to assign different weights to different parts of the input, processing information from different time steps more effectively. The basic structure of LSTM is shown in Fig. 2 (a), which includes an input gate, a forget gate, an output gate, cell state, and hidden state [35]. To improve the performance of the model, an attention mechanism based on LSTM units is used to reconstruct the TF of satellite cloud images. Its basic structure is shown in Fig. 2 (b).

Assume different TF sequences are represented by $X^k = \left(x_1^k, x_1^k, \cdots, x_L^k\right) \in R^L$. The AM is constructed based on the hidden state h_{t-1} and the cell state c_{t-1} of the LSTM unit:

$$\varepsilon_t^k = V_\varepsilon \tanh\left(W_{\varepsilon c} c_{t-q} + W_{\varepsilon h} h_{t-1} + b_i\right) \tag{12}$$

where, V_ε, $V_{\varepsilon c}$ and U_ε are parameters to be learned by the model. Then, the *softmax* function is applied to ε_t^k as the activation function. Which can be expressed as:

$$a_t^k = \frac{\exp\left(\varepsilon_t^k\right)}{\sum_{i=1}^{n}\exp\left(\varepsilon_t^i\right)} \tag{13}$$

Through the AM, the reconstructed satellite cloud image TF can be represented as:$\widetilde{x}_t = \left(a_t^1 x_t^1, a_t^2 x_t^2, \cdots, a_t^n x_t^n\right)$.

2.3. CNN-INF forecasting model

In the CNN-Informer model, a CNN layer is initially used to extract local features from the input data. These features are then fed into the self-attention layer of the Informer to capture the long-term dependencies within the sequence data. The self-attention mechanism is capable of assigning weights to different positions in the input sequence, effectively identifying and utilizing key information points, even if these points are far apart within the sequence. This approach combines the CNN's ability to detect local features with the Informer's capacity to recognize long-term relationships. The structure of CNN-INF as shown in Fig. 3(c).

2.3.1. The module of CNN

Firstly, the CNN layer is employed to extract local features from time series data. Take the input features as $X = [x_1, x_1, \cdots, x_T]$, where each x_t represents the feature input vector at time t. The CNN layer applies the following equation to perform feature extraction:

$$C_t = f(W_c * X_t + b_c) \tag{14}$$

where C_t represents the extracted features, W and b are the weights and biases of the convolution layer, * donates the convolution operation, and is the activation function ReLU.

2.3.2. The module of Informer

Subsequently, the output features from the CNN layer are fed into the Informer layer to capture long-term dependencies within the time series. The Informer accomplishes this through a self-attention mechanism, with its core formula as follows [36]:

$$A(Q,K,V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{15}$$

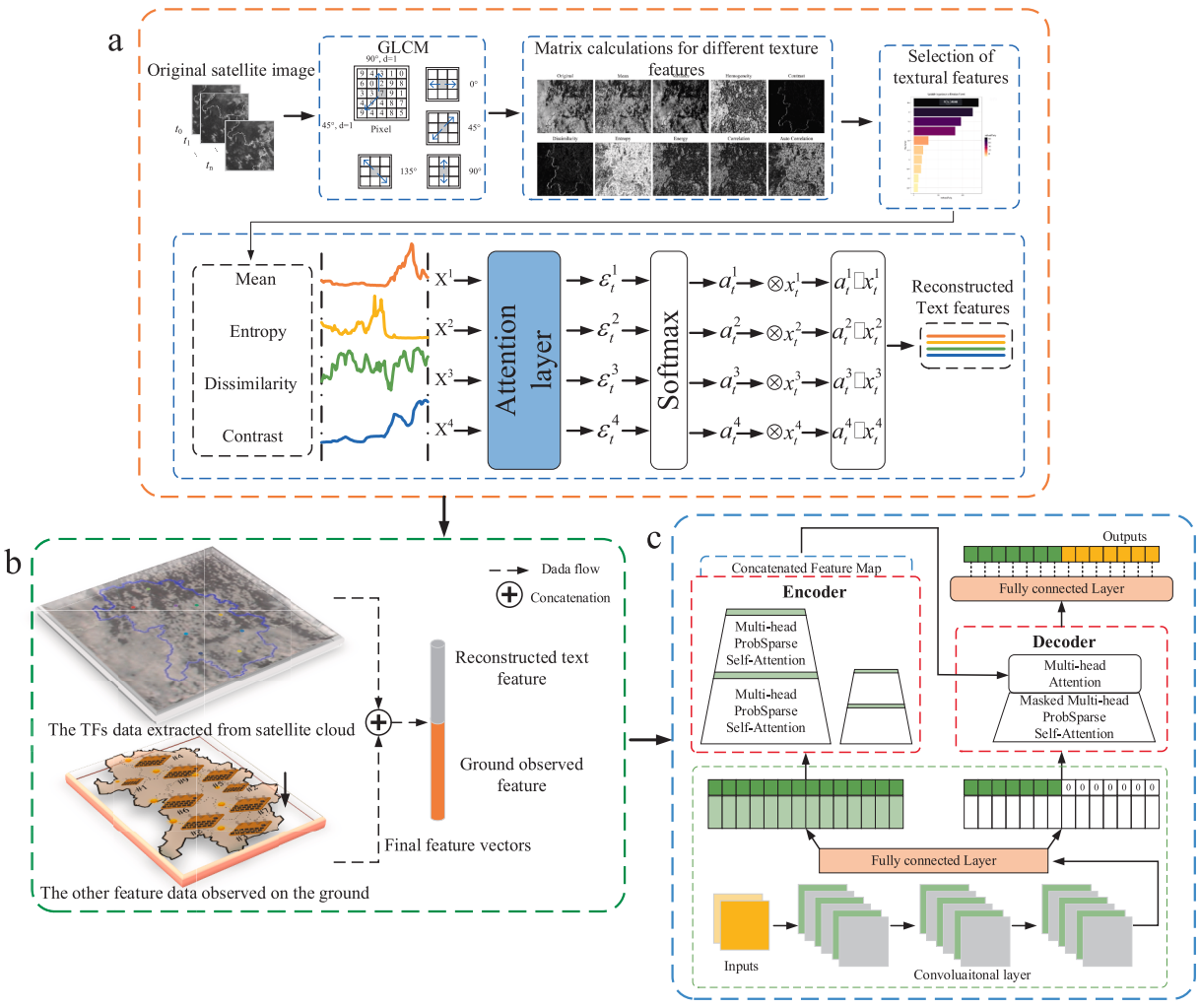


Fig. 3. The proposed framework. (a) Flowchart to extract and construct TF. (b) Feature concatenation. (c) The CNN-INF forecasting model.

where Q, K, and V represent the query, key, and value matrices, respectively, which are obtained by applying different linear transformations to the output features of the CNN. d_k represents the dimension of the key vectors, used to scale the dot products to enhance stability.

The softmax operation ensures that the sum of weights for each row equals 1. The primary function of the encoding layer is to filter out important information from the input feature sequence. The distillation operation applied in the encoder mainly aims to halve the length of the input sequence, thereby reducing the spatial complexity. Its calculation formula can be summarized as [37]:

$$X_{j+1}^{t} = MaxPool\left(ELU\left(Conv1d\left(\left[X_j^t\right]_{AB}\right)\right)\right) \tag{16}$$

where $[\bullet]_{AB}$ indicates key operations in the attention block, including multi-head ProbSparse self-attention, conv1d is a 1D convolution operation, ELU is the exponential linear unit activation function, and Maxpool denotes the max pooling layer.

This setup efficiently processes and emphasizes important features and patterns in the input sequence. The decoder layer of the Informer also utilizes a self-attention mechanism, which functions to generate the input sequence based on the information extracted by the encoding layer [38]. Its calculation formula can be summarized as:

$$X_{de} = \text{Concat}(X(token)X(0)) \in R^{\left(L_{token}+L_y\right)\times d_{model}} \tag{17}$$

where the $X(token) \in R^{L_{token}\times d_{model}}$ is the start token, and $X(0) \in R^{L_y\times d_{model}}$ is a placeholder for the target sequence, initialized with scaler values of 0. The output of the decoder layer is transformed into the final prediction output through a fully connected layer, which can be represented as [39]:

$$Y_{pred} = f\left(W_p H_{dec} + b_p\right) \tag{18}$$

where Y_{prep} represents the prediction values, H_{dec} represents the input features to the decoder, W_p and b_p are the weights and biases of the prediction layer, respectively.

2.3.3. *Framework of regional PV power forecasting*

The proposed methodology for forecasting power at regional PV stations is illustrated in Fig. 3, comprising three main stages: the extraction and reconstruction of TF from satellite cloud imagery based on their correlation with irradiance, feature concatenation, and the application of a CNN-INF model for training and forecasting. Initially, a sophisticated algorithmic framework that combines the GLCM and RF is utilized to extract and identify crucial TF from satellite images covering the regional PV stations. Subsequently, these TF, along with ground-observed features such as temperature, wind speed, and humidity, are amalgamated into final feature vectors. These concatenated feature vectors then serve as the input for the training of the CNN-INF forecasting model, optimizing it for accurate power prediction. This comprehensive process ensures a nuanced understanding of the factors influencing PV power generation, leading to more precise and reliable

forecasting outcomes.

The experiments in this study are conducted on a consistent platform. The software environment is built using TensorFlow and PyTorch within the Python programming language, while the hardware platform consists of an Intel Core i7-11700 CPU and an NVIDIA GeForce GTX 1660 Ti GPU. The architecture of the proposed CNN-INF forecasting model, along with the hyperparameter settings, is detailed in Table 1.

The hyperparameter settings are determined based on two experiments: single PV power forecasting and regional PV power forecasting. The convolutional layers are configured according to the structure of the training dataset, which is constructed using the proposed fusion feature that integrates the TF extracted from satellite images. For both the single PV and regional forecasting experiments, the pooling layer employs MaxPooling. This method adaptively extracts the maximum value from each channel without requiring a fixed input length, greatly enhancing application flexibility. Other hyperparameters, such as the learning rate, are set based on empirical experience to ensure that the characteristic data is thoroughly learned. The fully connected layer of the CNN is linked to the forecasting module of the Informer. The parameters of the Informer, including the hidden dimensions and the number of attention heads, are adjusted according to the specific horizon of the forecasting task. Additional hyperparameters are also established based on experience to optimize the learning of characteristic data.

2.4. Transfer learning strategy

In deep learning architectures, TL techniques have proven effective in addressing challenges such as suboptimal predictive accuracy due to data scarcity [40]. Additionally, they enhance computational efficiency by potentially reducing the training duration of models. With the increasing number of PV power stations in specific regions, relying on traditional forecasting models would result in significant computational time and resource consumption for predicting outputs across all stations. The TL strategy significantly accelerates the training speed of machine learning models by utilizing pre-trained models that have already captured general features from a large source dataset. This approach decreases the amount of training data required for the target task, allows for faster convergence during fine-tuning, and enables the transfer of useful representations across domains. Consequently, models can achieve optimal performance more rapidly, leading to lower computational costs and improved resource utilization, making transfer learning an efficient solution for adapting to new tasks. Hence, we employ a regional PV power forecasting model that leverages transfer learning and the proposed forecasting model.

Table 1
CNN-INF structure and parameter settings.

Structure	Hyperparameters Item	Single PV station	Regional PV stations
CNN module	Convolutional layer	Kernel size (3 × 3)	Kernel size (5 × 5)
	Activation function	ReLu	ReLu
	Pooling layer	MaxPooling (2 × 2)	MaxPooling (3 × 3)
	Fully connected layer	FC neurons (256)	FC neurons (512)
	Learning rate	0.001	0.002
Informer module	Number of layers	4	4
	Hidden dimension	64	64
	Number of attention heads	4	4
	Batch size	32	64
	Learning rate	0.001	0.002
	Regularization	L2	L2
	Dropout Rate	0.2	0.1

2.4.1. Selection of source and target domain PV station

In TL strategy, the foremost endeavor is the identification of optimally matched source and target domain datasets. In the study, the SOM [32] and MMD algorithms are utilized to get the source and target domain PV station. Firstly, the SOC is employed for the purpose of clustering in the regional PV power stations. It then utilizes the MMD algorithm to perform a detailed correlation analysis across each of the clustered PV stations. Leveraging the correlation metrics derived within each cluster, the research meticulously selects source and target domain PV stations that not only align with the predefined criteria for source domain data but also promise to enhance the efficacy of the transfer learning model.

The SOM algorithm is a simple neural network that consists only of an input layer and a competition layer. Neurons within the competition layer mutually enhance learning with their neighboring neurons until they possess similar weight vectors. The specific formula for the SOM algorithm can be referred to in the literature [32]. MMD is one of the most widely used loss functions in TL [41], especially in domain adaptation. It is mainly used to measure the distribution distance between two different but related random variables. In order to avoid the possibility of negative transfer [42], MMD is used to measure the differences between data from regional power stations, evaluated by a mean indicator. To avoid the potential for negative transfer, MMD is employed to measure differences in the data, and this discrepancy is assessed using a mean metric [43].

Given d_s (the source domain) and d_t (the target domain), the MMD distance between their distributions can be computed using a formula that measures the difference between sample means in a reproducing kernel Hilbert space. Specifically:

$$MMD(d_s, d_t) = \left\| \frac{1}{n_s} \sum_{i=1}^{n_s} x_{si} - \frac{1}{n_t} \sum_{j=n_s+1}^{n_s+n_t} x_{tj} \right\|_{\mathrm{H}} \tag{19}$$

where n_s, and n_t are the sample counts for the two domains respectively, x_{si} and x_{ij} are samples from the two domains respectively, $\| \bullet \|_H$ represents the Reproducing Kernel Hilbert Space (RKHS) norm [44]. The smaller the MMD value obtained by taking the difference in mean of the two sample distributions, the smaller the distribution difference, making it more suitable for TL. Within the domain of the target power stations, by calculating the source PV power stations with the smallest MMD value relative to that power station, the relevant data of that source PV power stations can be obtained.

2.4.2. Architecture of the CNN-INF based transfer learning method

The architecture of the CNN-INF-based TL model is illustrated in Fig. 4. The TL strategy not only accelerates the model's learning process but also reduces the volume of training data needed, showcasing its multifaceted superiority. In employing a TL approach for regional PV power forecasting, the initial step involves identifying source domain PV stations and target domain PV stations. The SOM and MMD algorithm are employed to identify the most correlated PV stations, with the station possessing the most comprehensive dataset designated as the source PV station within various groups.

This TL-based model encompasses two primary components: a CNN-based feature extraction module and an INF sequence forecasting module. In this TL configuration, a pre-trained CNN serves as the feature extractor, with its final fully-connected layer omitted, preserving only the convolutional and pooling layers. Subsequently, an INF layer succeeds the CNN module to facilitate sequence forecasting modeling.

The TL model is initially pre-trained on the source dataset to capture general feature representations [45]. Upon transfer to the target dataset, parameters in both the CNN and INF layers (excluding the fully connected layer) are fixed. And the parameters of fully connected layer are fined under retraining with the dataset of target PV stations. Subsequent fine-tuning occurs solely on the fully connected layer, allowing for task-specific adaptations. To mitigate overfitting, a reduced learning rate and

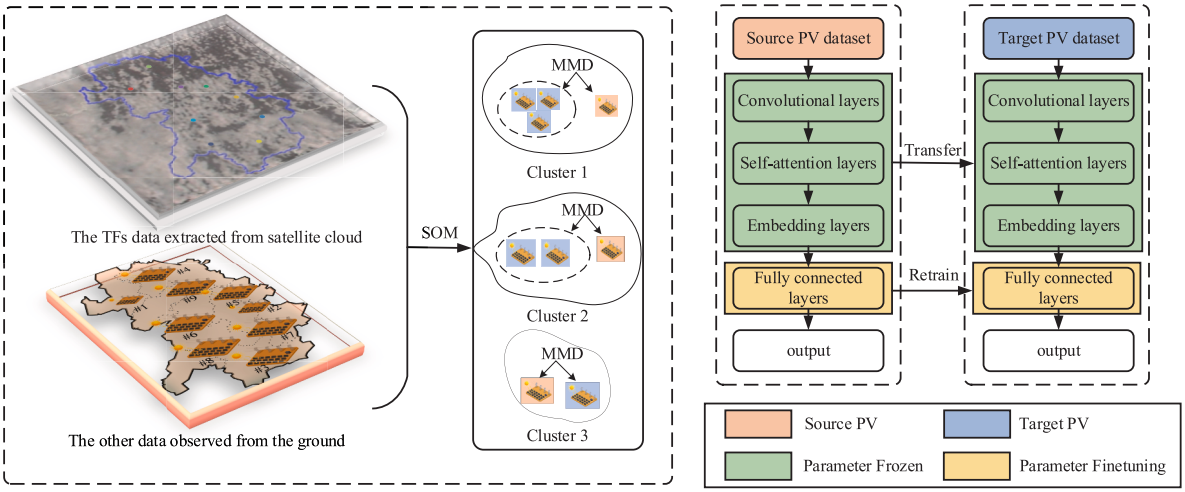


Fig. 4. Architecture of the TL model incorporating CNN-INF.

early stopping are implemented during this adjustment. Through iterative experimentation, optimal learning rates and other parameters are determined by selecting the model yielding the best training outcomes. A comprehensive summary of the training process is presented in Algorithm 1.

Algorithm 1 The training process of the TL architecture.

```
While not converged do
  Input data: source dataset
  Initialization: source dataset
    for training the source dataset do
    adjust parameters
    if e_MAE ≤ θ, and e_RMSE ≤ τ
      lock and transfer parameters in CNN and INF layers
      end
    end
  Input data: target dataset
  Initialization: target dataset
    for training the target source dataset X′ do
      finetune FC parameters
      if e′_MAE ≤ θ, and e′_RMSE ≤ τ
      lock FC parameters
      end
    end
end
```

where,X and X' are the input data sets for the source domain PV station and target domain PV station, respectively. e_{MAE} and e_{RMSE} denote the predicted values of MAE and RMSE in the source domain PV station, respectively. e'_{MAE} and e'_{RMSE} represent the MAE and RMSE values in the target domain PV station. θ and τ thresholds for the values of MAE and RMSE, respectively.

3. Case study

In this section, two primary experiments were carried out: one focusing on single PV power forecasting, and the other on regional PV power forecasting. The experiment on the single PV station aims to verify that the constructed TF and the CNN-INF model can substantially improve forecasting accuracy. On the other hand, the experiment on regional PV stations aims to demonstrate that the transfer strategy can significantly boost computational efficiency.

3.1. Data description

3.1.1. Data from satellite cloud images

The satellite cloud imagery data are sourced from the public meteorological satellite Fengyun-4 (https://rsapp.nsmc.org.cn/) [46]. Continuous image datasets are collected with a maximum temporal resolution of 5 min and a spatial resolution of 0.25 km. This imagery meets the spatial and temporal accuracy requirements for the PV station study. On the associated public data platform, users have the option to select and export satellite cloud imagery specific to designated longitude and latitude regions. The exported imagery is saved in GIF format, and a maximum of 6 months' historical image datasets can be collected from this platform. An example of a satellite cloud image at a specific time step is shown in Fig. 5.

3.1.2. Data from single PV station

The data for the single PV station are collected from Brunel University London in the UK. The PV installation is located atop a campus building and is monitored by the comprehensive Sunny Portal system, which logs detailed information on irradiance, temperature, wind speed, and power output. An overview of the PV installation is shown in Fig. 6. For experimental purposes, we focus on PV data from October 1 to October 14, 2022, as a representative sample. Fig. 7 presents the PV power data for this interval, with data showing zero PV power during nighttime hours excluded to emphasize the temporal variations in PV power.

3.1.3. Data from regional PV stations

The data from regional PV stations are collected from a specific area in China, which includes irradiance data from nine PV stations. The geographical distribution map of these regional PV stations, marked by the latitude and longitude of each PV site as shown in the satellite cloud image, is presented in Fig. 8.

Furthermore, the key information for each PV station, including capacity, altitude, latitude, longitude, tilt angle, and the number of

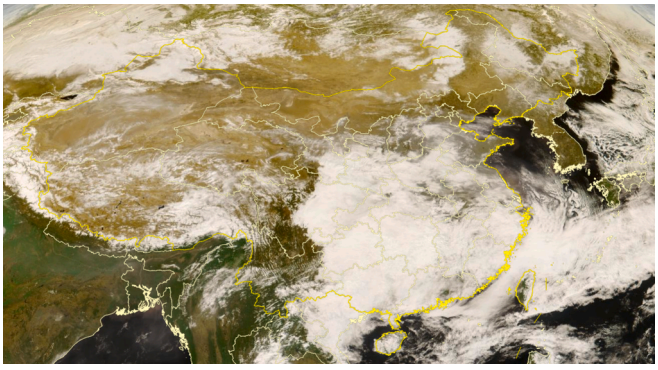


Fig. 5. Overall satellite cloud imagery captured at a specific time step.

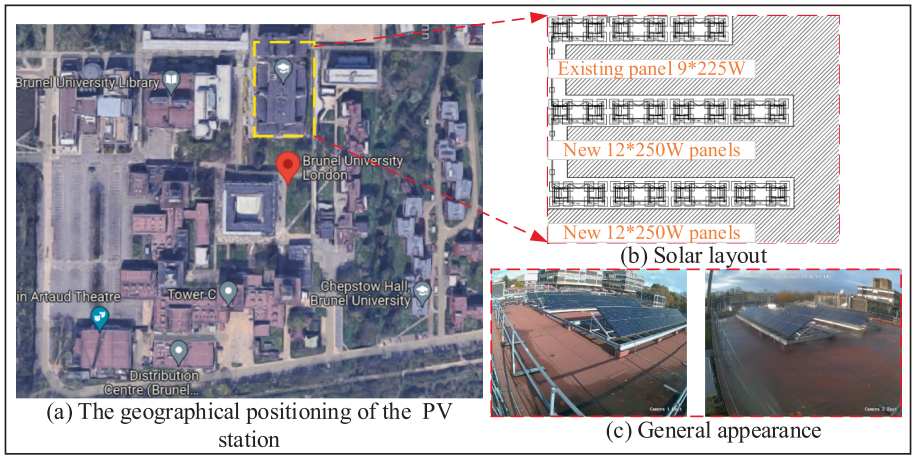


Fig. 6. Overview of a single PV station in the UK.

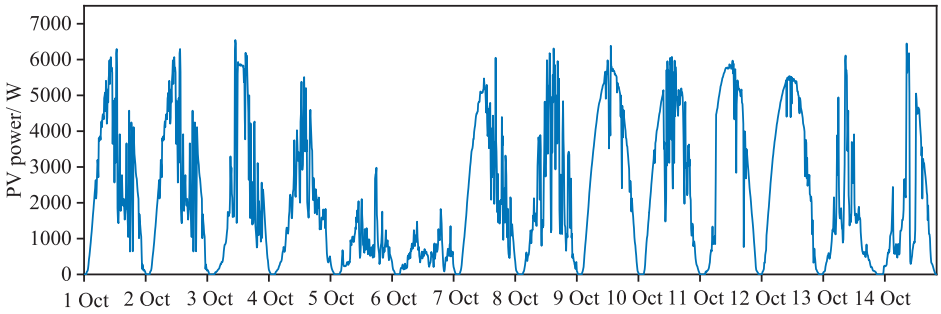


Fig. 7. Two-week PV power curve from a UK station.

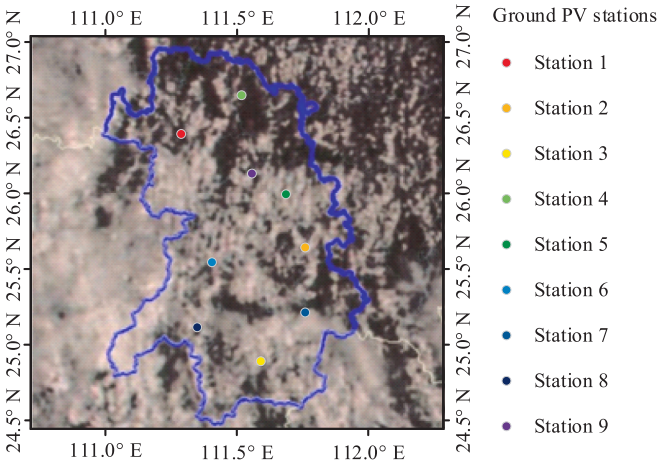


Fig. 8. Geographical distribution map of regional PV stations.

inverters, is provided in Table 2.

This dataset, which includes timestamps, irradiance metrics, and historical power figures, was provided by the China Grid Company and can be accessed at https://github.com/xieyang0929/regional-PV-dataset. For our study, we selected power readings under various weather conditions from October 1 to October 14, 2022, to serve as the forecasting target. Fig. 9 showcases the power curves from these PV stations during the specified two-week interval.

3.2. Evaluation metrics

The primary objective of this research is to enhance both forecasting accuracy and computational efficiency for regional PV power stations. Consequently, evaluation metrics are derived from two main facets: computation time and forecasting accuracy. The computation time accounts for forecasting across a variable number of PV stations, all while maintaining consistent forecasting accuracy. For gauging forecasting accuracy, we utilize standard statistical metrics: mean absolute error (MAE), mean bias error (MBE), root mean square error (RMSE), and R-squared. The relevant formulas for these metrics can be described as follows [47]:

Table 2
Key information of the regional stations.

PV stations	Latitude and longitude	Capacity (kW)	Altitude (m)	Tilt angle (°)	Number of inverters
No. 1	(E 111°18′0″, N16°37′0″)	10.0	259	35	6
No. 2	(E111°45′0″, N25°39′57″)	12.0	287	35	8
No. 3	(E111°36′0″, N24°45′0″)	10.0	325	36	6
No. 4	(E111°30′0″, N26°37′30″)	8.0	345	36	5
No. 5	(E111°40′1″, N26°0′0″)	8.0	257	37	5
No. 6	(E111°24′0″, N25°30′0″)	8.2	335	35	6
No. 7	(E111°45′0″, N25°15′0″)	9.0	405	36	6
No. 8	(E111°20′24″, N25°6′0″)	9.0	320	38	6
No. 9	(E111°36′0″, N26°6′0″)	10.0	243	37	6

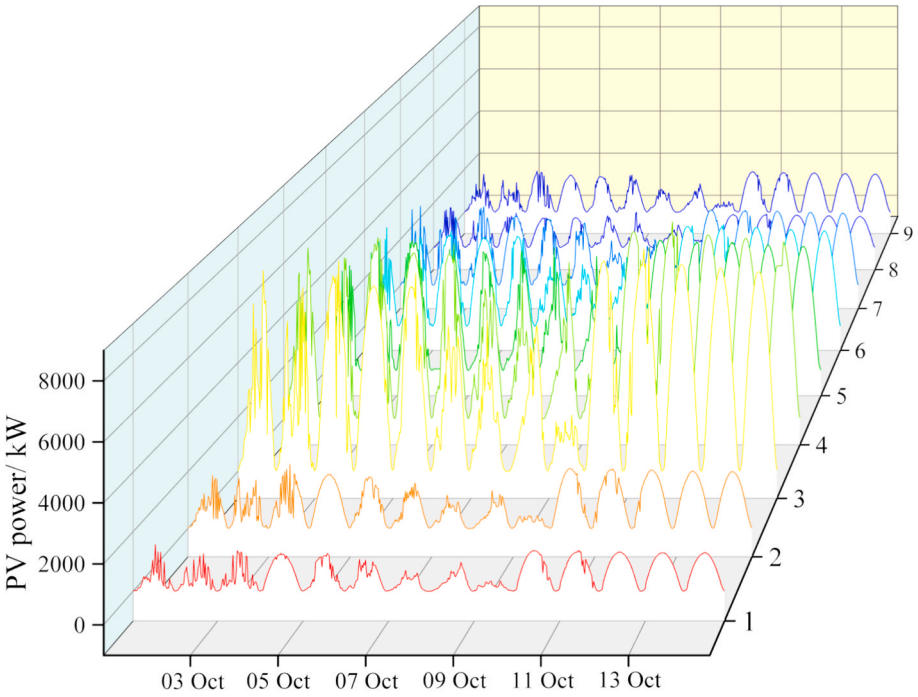


Fig. 9. Two-week PV power curves from nine stations in a Chinese region.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(y(i) - y_p(i)^2\right)}{\sum_{i=1}^{N}(y(i) - \bar{y})^2} \tag{20}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y(i) - y_p(i)\right)^2} \tag{21}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y(i) - y_p(i)\right| \tag{22}$$

$$\mathrm{MBE} = \frac{1}{N}\sum_{i=1}^{N}\left(y(i) - y_p(i)\right) \tag{23}$$

where $y(i)$ and $y_p(i)$ represent the actual PV power and predicted PV power, respectively, and N denotes the number of sample points. To evaluate the efficacy of the comparative model in tracking dynamic cloud changes, we employ the Persistence Forecast Model (PFM) as a benchmark for forecasting skill. This approach is inspired by methodologies detailed in reference [47]. The clearness-based PFM is formulated as follow:

$$y_p(i+1) = K(i) \times y_e(i+1), K(i) = y(i) \Big/ y_e(i) \tag{24}$$

where y_e represents the value of PV power corresponding to extraterrestrial solar radiation, while K is the clearness index, serving as an approximate indicator of cloud transmittance. Then, the metric of forecast skill can be express as:

$$\text{Forecast skill} = 1 - \frac{\mathrm{RMSE}}{\mathrm{RMSE}_{PFM}} \tag{25}$$

Additionally, given the sequential characteristic of PV power forecasting, maintaining phase consistency in time series forecasting is crucial for effective model evaluation. To gauge the synchronicity of the forecast model, we employ the lag ratio [47], a measure designed to assess the phase lag in the forecasted time series. The mathematical formulation of this concept is outlined below:

$$\text{Lag ratio} = 1 - \frac{1}{N-1}\sum_{i=1}^{N-1}\sigma\Big(sign\Big(y_p(i+1) - y_p(i)\Big) - sign(y(i+1) - y(i))\Big) \tag{26}$$

where $sign(\bullet)$ donates the sign function, and $\sigma(\bullet)$ donates the sigmoid function. Therefore, the effectiveness of the various models is assessed using a range of statistical measures, including MAE, MBE, RMSE, R^2, forecast skill, and the lag ratio.

3.3. Calculation and selection of the TF

3.3.1. Result of TF calculation

Before calculating the TF of the cloud image, the original satellite cloud image needs to be converted to grayscale. In this study, the average method is used. Then, the method of GLCM is employed to calculate the TF of satellite cloud image.

Time-series satellite cloud images undergo consistent processing to convert to grayscale, ensuring timestamps are retained for each image. Following the grayscale conversion, the GLCM is computed for each image, serving as a foundation for TF calculation. Parameters are set with a gray level of 64, a sliding window size of 7, a stride of 2, and an averaged direction from 0°, 45°, 90°, and 135°. Notably, the original image holds a gray level of 256 with dimensions of 512*512. Fig. 1. illustrates the TF calculation process based on the GLCM.

After calculation, the results of different TF are shown in Fig. 10. The figure includes the original satellite cloud image, and features such as mean, variance, homogeneity, contrast, dissimilarity, entropy, energy, correlation, and auto correlation.

3.3.2. Selection of the important TF

Before analyzing the impact of different TF on irradiance, due to the numerous TF of satellite cloud images, using all TF as features of cloud will inevitably cause feature redundancy and increase the computational load of the model. Therefore, the RF algorithm is used to rank and select the importance of different TF.

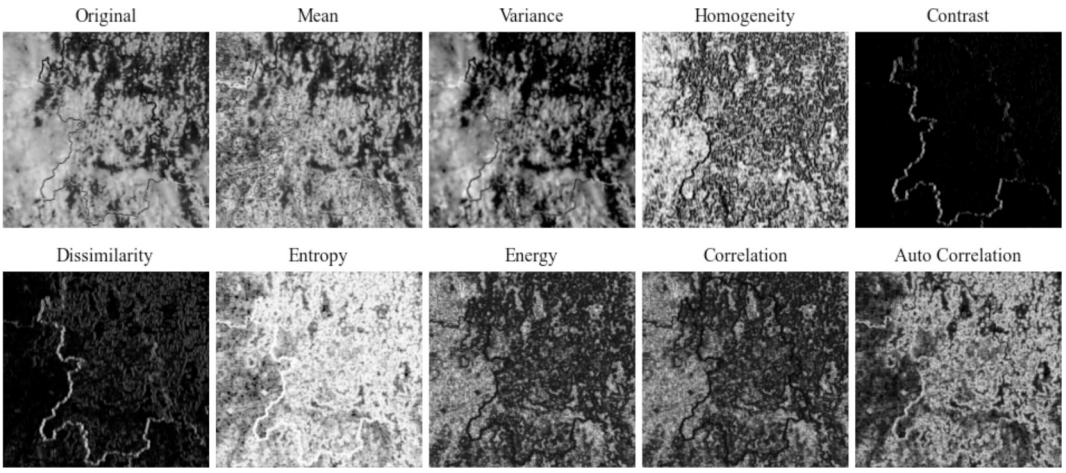


Fig. 10. Results of different TF calculations.

Before using RF to calculate the importance of different TF, it is necessary to preprocess the obtained TF. First, merge the TF variables of the same time series with irradiance variables. In the new data structure, the first column is set as irradiance data, and the rest are TF data. Thus, the first column is referred to as the target variable Y in the RF, and all columns after the first column are used as features. Then, use 80 % of this dataset as the training set and 20 % as the test set. The number of decision trees is set to 100, and the random_state is set to 42, indicating the chosen seed value. In the RF, the importance of features can be calculated by measuring the degree of reduction in node purity for each feature when constructing decision trees. Utilizing this approach, Fig. 11 (a) displays the importance ranking of various TF for irradiance, while Fig. 11 (b) presents the PDPs of different TF.

3.4. Verification of the proposed method on general performance

The purpose of this section is to validate the general performance of the proposed method, particularly the effectiveness of TF and the forecasting model for PV power forecasting. This includes assessing forecasting accuracy under various conditions, such as differing weather patterns, multiple horizons, and overall robustness. The special performance of regional PV power forecasting will be validated in the next section, where the number of PV sites is increased. Therefore, this section uses data provided by Brunel University London in the UK for training and verification.

To gauge the efficacy of our approach, we consider a multifaceted validation strategy. Firstly, we account for varied weather conditions, understanding that cloud layers display distinct patterns under different meteorological states. Specifically, we've chosen test validation sets from dates capturing diverse conditions: scattered clouds (October 1, 2022), broken clouds (October 3, 2022), partly clouds (October 4, 2022), and drizzle (October 5, 2022). Moreover, considering the varying data volume and utilization rates across PV stations, the model's robustness is crucial. As such, we've incorporated experimental validation to underscore the model's generalization and robustness.

3.4.1. Model generalization validation

After obtaining the cloud TF at the location of the photovoltaic power station, it is recognized that the variations in cloud conditions under different weather scenarios can lead to corresponding changes in the power output of the photovoltaic station. To verify the effectiveness of the method adopted in this paper, under the premise of the same amount of training data, various methods are employed to validate the model. The comparison methodologies are outlined as follows:

- NWP-CNN-LSTM: The method dubbed "NWP-CNN-LSTM" utilizes NWP data to train a combined CNN and LSTM model for forecasting purposes.
- TF-CNN-LSTM: The method dubbed "TF-CNN-LSTM" utilizes raw TF as part of input data to train a combined CNN and LSTM model for forecasting purposes.
- TF-AM-LSTM: This approach incorporates TF reconstructed using an LSTM-based AM as part of input data and employs an LSTM network for predictive modeling.
- TF-AM-CNN-INF (proposed): This approach incorporates TF reconstructed using an LSTM-based AM as part of input data and employs the model combined CNN and INF for predictive modeling.

The validation is performed using power data under various weather conditions from the region over a week, specifically from October 8th, 2022 to October 14th, 2022. The weather variations during this period are depicted in Fig. 12, and the forecasting result under vaious weather conditions are present in Fig. 13.

To demonstrate the model's capability to capture and compute various types of cloud changes, Fig, 13(a)-(d) respectively show the PV power forecasting under the conditions of scattered cloud, broken cloud, and passing cloud. The corresponding dates for these conditions are October 13th, October 12th, and October 11th. (8 h ahead forecasting).

From the aforementioned forecasting results, it can be observed that the main source of error in different forecasting models comes from intervals with significant power fluctuation rates. When correlated with the weather data during this period, it becomes evident that the abrupt power changes are primarily caused by weather variations, with cloud cover being a major factor. Therefore, the TF-AM-CNN-INF forecasting model, which excels in capturing and computing cloud cover characteristics, has the smallest power forecasting error.

Furthermore, to validate the advantages of the innovative TF data and the TL-based forecasting model, the proposed method is compared with another four additional fundamental forecasting algorithms. These include the PFM, extreme gradient boosting (XGBoost), sequence to sequence (seq2seq), and gated recurrent unit (GRU), respectively. Eight-hour ahead PV power forecast accuracy on testing sets under different weather condition are shown in Fig. 14.

It is important to highlight that the methodologies incorporating TF utilize concatenated feature vectors as training data. In contrast, alternative approaches rely solely on feature vectors observed on the ground as their training data. This distinction underscores the added value of TF in enriching the training dataset, potentially leading to enhanced forecasting performance by providing a more comprehensive data foundation. Notably, the TF-AM-CNN-INF model consistently outperforms others in diverse meteorological conditions, ranging from clear skies to

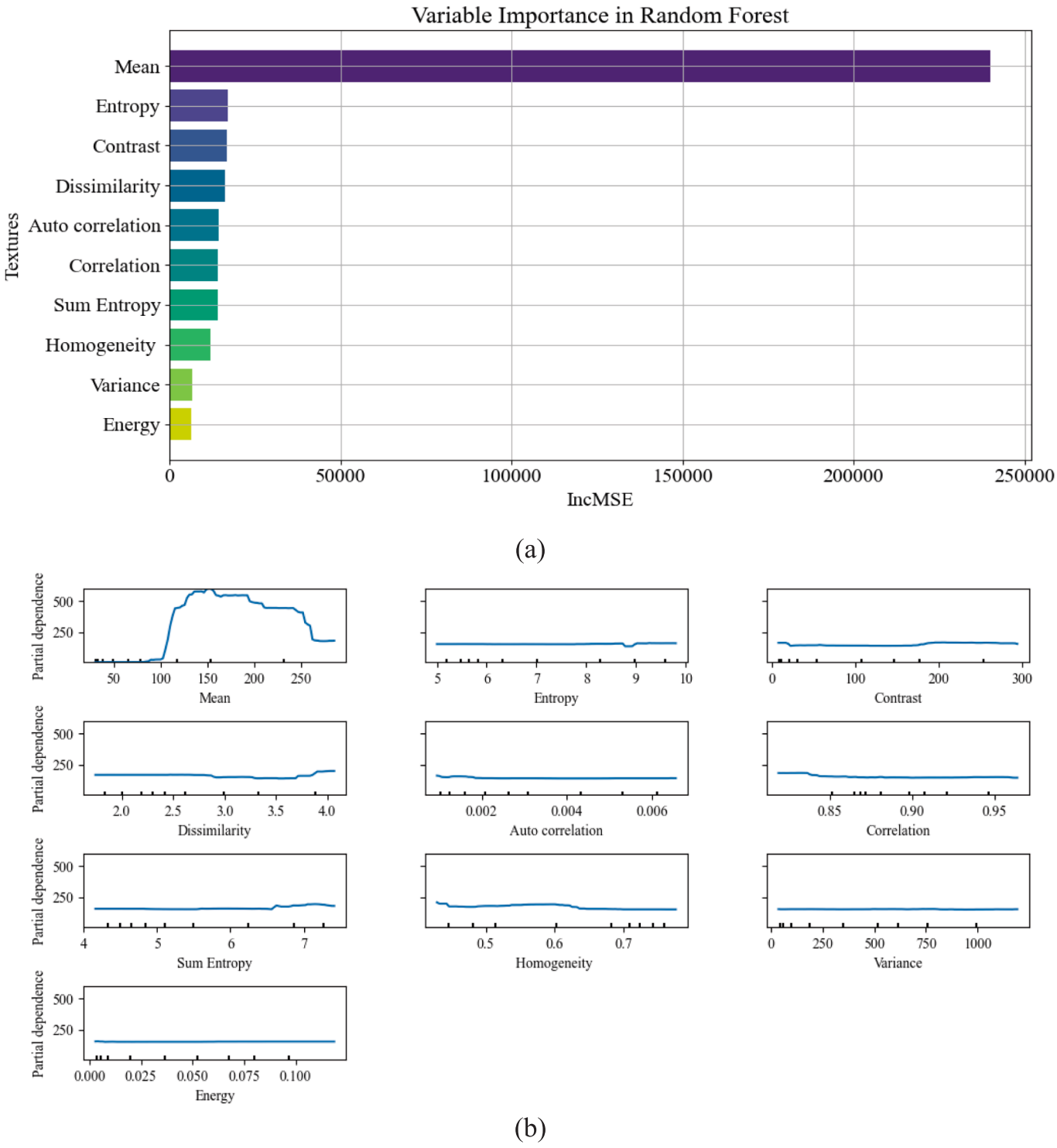


Fig. 11. Ranking of TF importance using RF algorithm. (a) Variable importance ranking. (b) PDP curve.

those with significant cloud variations. On the contrary, TF_CNN_LSTM's forecasts are the worst among the methods with TF because reconstructed TF via AM can capture the motion of the cloud better than original TF. Compared with CNN_LSTM, the maximum improvement of TF-AM-CNN_INF is approximately 8.7 % in term of RMSE and 2.5 % in term of lag ratio on average. Compared to models without TF training data, those incorporating TF effectively capture the dynamic nature of cloud movements. Consequently, in cloudy weather conditions, models trained with TF data exhibit superior performance in weather with cloud variations compared to clear weather conditions. Notably, the TF-AM-CNN-INF model, proposed in this paper, has the smallest forecasting error. This can be attributed to the model not only using cloud TF as part of the training data but also incorporating the attention mechanism to reconstruct different TF, allowing for precise dynamic capturing of power variations caused by cloud changes. This superior performance is largely due to the synergetic integration of CNN's robust feature extraction capabilities and INF's proficiency in processing time series data, rendering it exceptionally suitable for spatial–temporal data analysis.

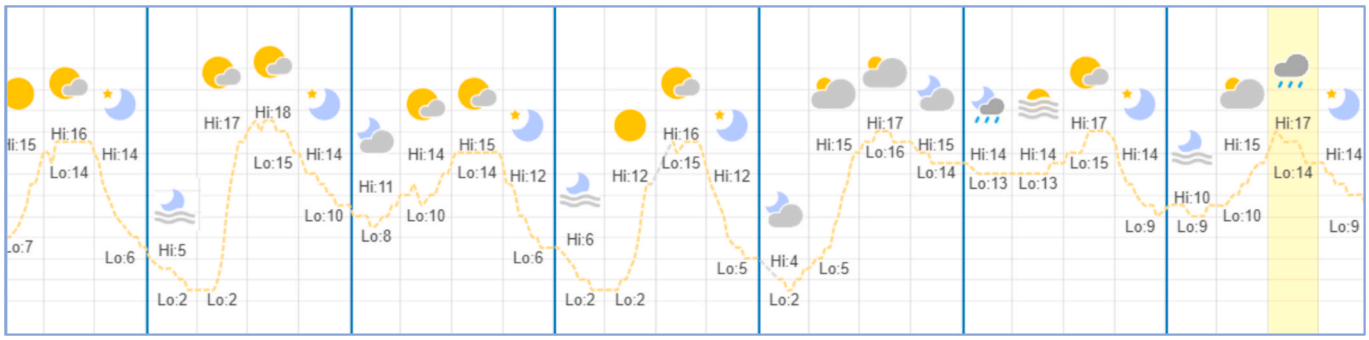


Fig. 12. Weather condition during the validation period.

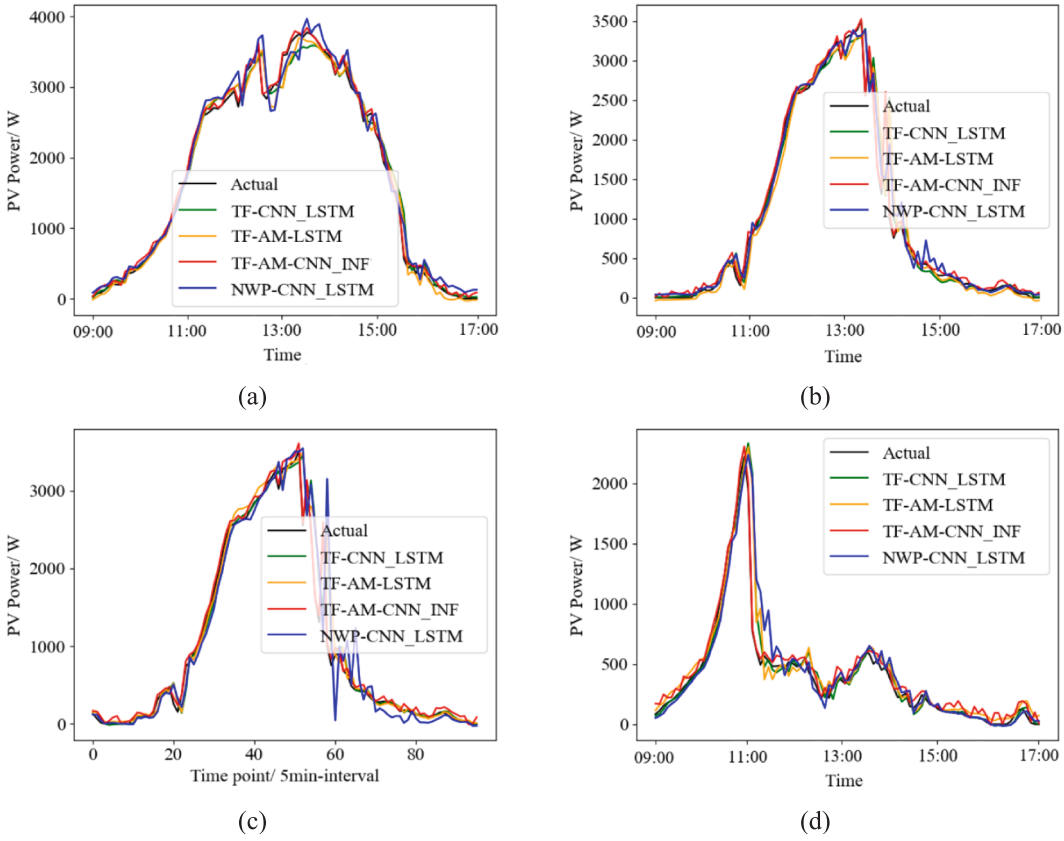


Fig. 13. Forecasting results for different clouds weather conditions. (a) sunny day (b) scattered clouds. (c) broken clouds. (d) passing clouds.

3.4.2. Forecasting performances over multiple horizons

The TF extracted from satellite images have been shown to significantly enhance forecasting accuracy, indicating an improvement in the spatial resolution of PV power forecasting. Additionally, the temporal resolution of PV power forecasting is also crucial for modern power systems. In this section, the forecasting performance across multiple horizons will be validated by comparing state-of-the-art methods.

The forecasting performance is influenced not only by the feature data but also by the forecasting model used. The LSTM-based AM is designed to improve the temporal resolution of the feature data. The comparison benchmarks will include the method trained with TF feature data without the AM, as well as the CNN-INF model. Additionally, to highlight the significance of the TF, we will also compare it against a method that utilizes traditional feature data with the CNN-INF forecasting model. Furthermore, state-of-the-art models based on TF extracted from satellite images will be included as comparison benchmarks to demonstrate the superiority of the proposed method in this research. The comparison of the forecasting methods can be summarized as follows.

Method 1 (CNN-INF): Traditional feature data—such as historical PV power, temperature, wind speed, and others—are used without the TF and forecasted with the CNN-INF model.

Method 2 (TF-CNN-INF): The TF extracted from satellite images are utilized as additional feature data and forecasted using the hybrid CNN-INF model.

Method 3 (TF-PCPOW): The TF extracted from satellite images is used as additional feature data for forecasting with the model PCPPW proposed in the research [18].

Method 4 (TF-SVM-INF): The TF extracted from satellite images is used as additional feature data for forecasting with the model SCM-INF proposed in the research [19].

Method 5 (TF-AM-CNN-INF): The TF extracted from satellite images is reconstructed using an LSTM-based attention mechanism before being forecasted by the CNN-INF model.

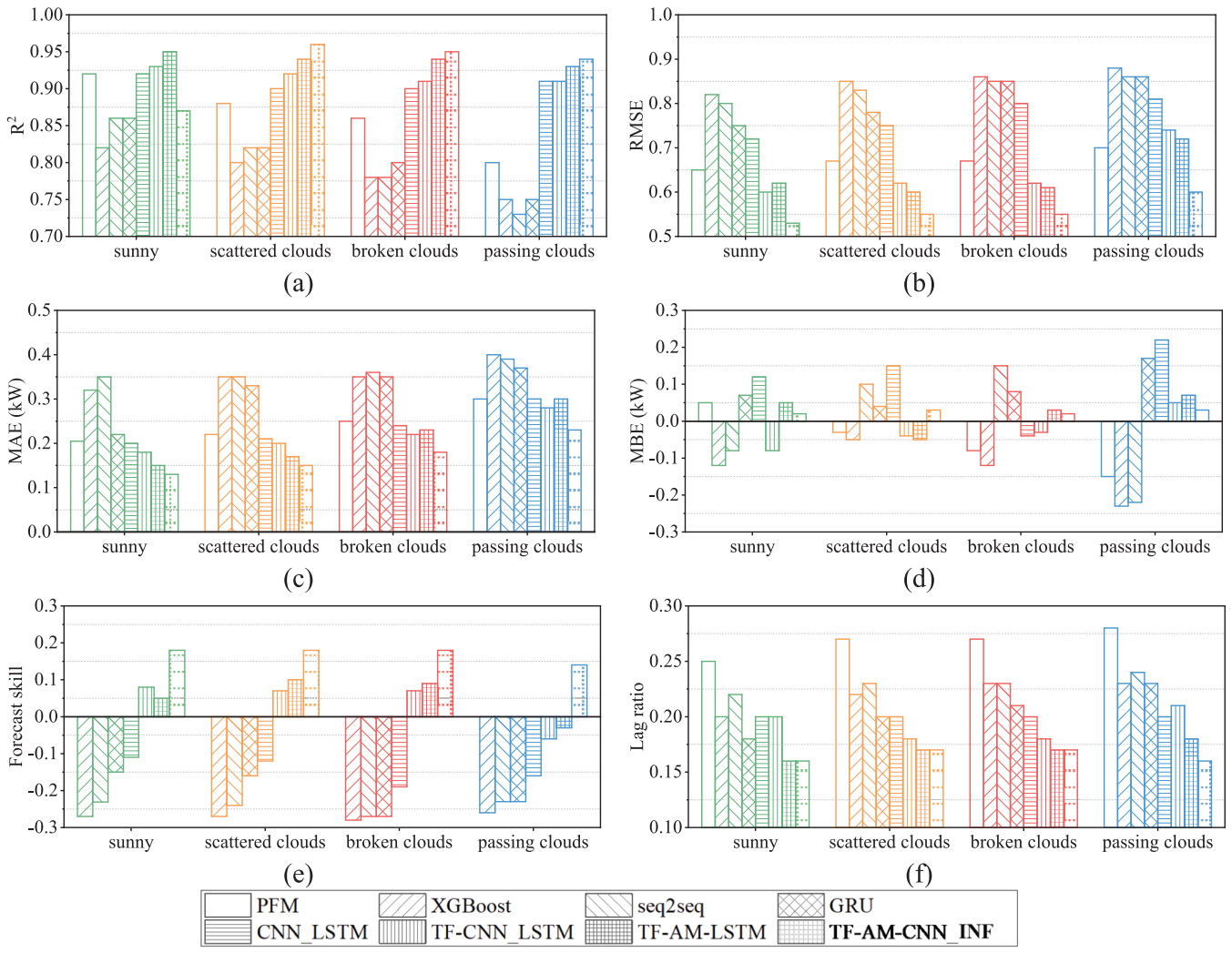


Fig. 14. Eight-hour ahead PV power forecast accuracy on testing sets under different weather condition. (a) R^2. (b) RMSE. (c) MAE. (d) MBE. (e) Forecast skill. (f) Lag ratio.

The forecasting performance is compared across different methods over multiple horizons, which include 10 min, 30 min, 1 h, 2 h, 3 h, 4 h, 5 h, 6 h, 7 h, and 8 h. The forecasting results are summarized in Table 3.

To more comprehensively demonstrate the forecasting performance across various models, evaluation metrics for multiple time horizons have been statistically summarized and are presented in Fig. 15.

It is clear that the Method 5 proposed in this study achieves the best forecasting performance across multiple horizons. In contrast, Method 1 exhibits the poorest performance, as the feature data was trained without the TF extracted from satellite images, leading to an increased lag ratio. Method 2 outperforms Method 1 due to the inclusion of additional TF data, which provides valuable information and insights into cloud motion correlations. When comparing Method 2 with Method 5, the latter demonstrates stable and superior performance. This advantage is primarily due to the attention mechanism introduced in this research, which significantly enhances the model's ability to capture temporal correlations and improve the temporal resolution of the feature map. Methods 3 and 4 show similar performance across various horizons but perform worse than the proposed Method 5. This is because the correlation within the feature map constructed with TF extracted

Table 3
The forecasting results across multiple horizons.

Method	Metrics	Forecasting horizon									
		10 min	30 min	1 h	2 h	3 h	4 h	5 h	6 h	7 h	8 h
Method 1	RMSE	0.035	0.048	0.058	0.072	0.085	0.115	0.143	0.187	0.232	0.279
	Lag ratio	0.13	0.15	0.20	0.24	0.25	0.27	0.28	0.28	0.29	0.31
Method 2	RMSE	0.012	0.013	0.021	0.032	0.047	0.069	0.094	0.104	0.135	0.173
	Lag ratio	0.10	0.12	0.13	0.16	0.16	0.18	0.18	0.23	0.27	0.28
Method 3	RMSE	0.014	0.015	0.015	0.024	0.031	0.054	0.072	0.103	0.112	0.131
	Lag ratio	0.11	0.12	0.15	0.19	0.21	0.22	0.25	0.25	0.26	0.26
Method 4	RMSE	0.014	0.015	0.018	0.019	0.022	0.031	0.042	0.051	0.064	0.104
	Lag ratio	0.12	0.13	0.13	0.15	0.16	0.17	0.17	0.20	0.24	0.25
Method 5	RMSE	0.015	0.016	0.017	0.018	0.020	0.021	0.022	0.022	0.024	0.025
	Lag ratio	0.12	0.12	0.13	0.14	0.14	0.14	0.14	0.15	0.15	0.16

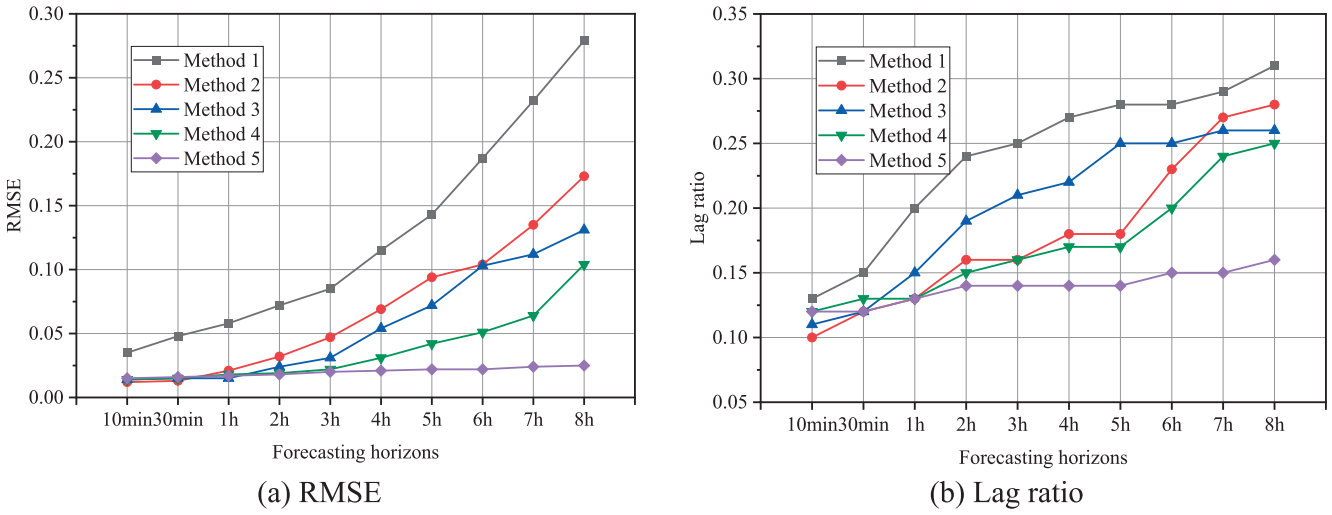


Fig. 15. Overall performance of different models across multiple time horizons, presented through a comparison of (a) RMSE and (b) Lag Ratio.

from satellite images is more effectively captured by the CNN module of the proposed forecasting model. It should be noted that the forecasting performance of the proposed method declines as the forecasting horizon increases. This decline is attributed to the constraints of the satellite image dataset obtained from public platforms, as well as the challenges in capturing the long-term dynamic features of the satellite images using the proposed method.

3.4.3. Model robustness validation

Same models can present different forecasting performances in different training data, especially in the power forecasting of PV stations. This is because in PV stations, there are differences in data sampling frequency and the number of samples. Therefore, the robustness of the model is also an important indicator to test its performance.

To verify the robustness and anti-interference ability of the model, 5-fold cross-validation is used to validate the forecasting results [48]. Specifically, this dataset is divided into 5 different subsets. In each round of validation, one of the 5 subsets are designated as the validation set, while the remaining 4 subsets form the training set. Then, the model is trained using the training set and evaluated using the validation set. Performance evaluation is done using metrics such as MSE and forecasting accuracy, and these evaluations are carried out after each round of validation. This generates five sets of performance evaluation results. This facilitates the analysis of the stability and generalization ability of various forecasting models. Since robustness tests the performance of the model itself and is unrelated to the trained data, all comparison models have the same data. The comparison models include the LSTM model, LSTM forecasting model based on the attention mechanism, CNN-LSTM, and the AM-CNN-INF model proposed in this paper. The result of 5-fold cross-validation are presented in Table 4.

From the table, it is evident that the CNN-INF forecasting model, which utilizes the attention mechanism proposed in this article, demonstrates strong robustness. This robustness is attributed to the enhanced generalization capabilities provided by the attention mechanism, which effectively captures the correlations within the input feature dataset. Furthermore, the encoder and decoder of the Informer, structured in a multi-layer stacked design, allow the model to learn more complex features and patterns. This stacked architecture aids in capturing information at various levels, thereby improving the model's adaptability to changes in the input data.

Table 4
MSE value and accuracy of different forecasting models in cross-validation.

Fold	Index	LSTM	AM-LSTM	CNN-LSTM	AM-CNN-INF
1	MSE	0.128	0.0285	0.0236	0.0135
2		0.146	0.0357	0.0277	0.0145
3		0.167	0.0255	0.0215	0.0112
4		0.175	0.0415	0.0271	0.0139
5		0.195	0.0255	0.0224	0.0128
Avg.		0.1622	0.03134	0.02446	0.01318
1	Accuracy	0.798	0.927	0.935	0.987
2		0.766	0.903	0.928	0.935
3		0.811	0.918	0.928	0.977
4		0.808	0.911	0.915	0.929
5		0.783	0.917	0.929	0.946
Avg.		0.7932	0.9152	0.927	0.9548

3.5. Verification of the proposed method on special performance

As the number of PV sites increases, the validation of special forecasting performance will be addressed in this section. This proposed method, particularly the transfer learning strategy developed in this research, aims to alleviate the computational burden, especially in regional PV power forecasting.

3.5.1. Correlation analysis between different PV power stations

Before employing a TL model for power forecasting of regional PV stations, it is necessary to analyze which PV power station within the domain has features most similar to the current PV power station. This aids in training the model to obtain more appropriate model parameters. To determine the source and target PV stations for our analysis, we first obtain the results from the SOM clustering, as outlined in Table 5.

Then, the selection of the source and target PV station is based on the magnitude of correlations among photovoltaic (PV) stations within each cluster. Equation (19) provides the theoretical calculation method for MMD. In actual computer simulation, it needs to be processed in matrix form. The transformation process of this equation can be expressed as follows [49]:

Table 5
SOM clustering results.

Category	PV station number
A	4, 8, 9
B	3, 6, 7, 8
C	1, 2

$$MMD^2 = \left\| \frac{1}{n}\sum_{n}^{i=1} f(x_i) - \frac{1}{m}\sum_{m}^{j=1} f(x_j) \right\|_{\mathscr{H}}^2$$
$$= \left\| \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{n} k(x_i, x_i') - \frac{2}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} k(x_i, y_j) + \frac{1}{m^2}\sum_{j=1}^{m}\sum_{j'=1}^{m} k(y_j, y_j') \right\| \quad (27)$$
$$= tr\left(\begin{bmatrix} K_{s,s} & K_{s,t} \\ K_{t,s} & K_{t,t} \end{bmatrix} \boldsymbol{M} \right)$$

where **M** can be expressed as:

$$(M)_{ij} = \begin{cases} \frac{1}{n_s n_s}, \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathscr{D}_s \\ \frac{1}{n_t n_t}, \boldsymbol{x}_i, \boldsymbol{x}_j \in \mathscr{D}_t \\ \frac{-1}{n_s n_t}, \text{ otherwise} \end{cases} \quad (28)$$

where k represents the kernel matrix obtained from the Gaussian kernel function, which maps each sample via the Gaussian kernel function. By utilizing the PyTorch framework in Python for programming calculations, we determined the MMD values between any two photovoltaic power stations in this region. Using the Gaussian kernel function, with a bandwidth set to 1, the results are shown in Fig. 16.

It can be observed that the MMD values between different PV stations present a symmetrical matrix structure. In the symmetrical matrix, elements are symmetric along the main diagonal. From this figure, one can easily determine the PV stations that have the most similar meteorological-power features with another, that is, the minimum MMD value in the corresponding column. Specifically, the station most similar to other stations in terms of meteorological-power features is station No.4 in cluster A. Similarly, the station most similar to other stations is station No.6 in cluster B.

3.5.2. Forecasting performance under different composition of training data

The primary objective of this part of experiment is to validate the efficacy of the proposed model not only in source domain PV stations but also in target domain stations. Furthermore, the reconstructed TF play a pivotal role in enhancing the model's performance. Then, the comparing methods are designed as follows:

- TL: This method exclusively utilizes NWP data as input and employs a forecasting model based on TL as conceptualized in this study.
- TF-TL: This method utilizes the raw TF as part of input data and employs a forecasting model based on TL as conceptualized in this study.
- TF-AM-TL (proposed): This approach utilizes the constructed TF as part of input data and employs a forecasting model based on TL as conceptualized in this study.

To verify the generalization of the method proposed in this study, data under different weather conditions from the validation dataset have been selected for forecasting analysis. The forecasting results for the source domain PV station and target domain PV station under different weather conditions are compared. Taking the PV station in cluster A as an example, the forecasting results are shown in Fig. 17.

From Fig. 17(a) and (b), it is evident that the forecasting model based only on TL had the highest error rate. This is because, under cloudy conditions, the variations in cloud layers significantly impact PV power output. Solely relying on the TL model for forecasting neglects the effect of cloud layers. Additionally, while the predictive results based on the TL model and take the raw TF as part of input data, they were still slightly inferior to the model proposed in this study, which combines TL, satellite cloud imagery, and the AM. This superiority is attributed to the incorporation of the AM, which assigns different weights to various TF in the input data, enabling more precise capture of power fluctuations due to cloud variations at different times.

From Fig. 17(c)-(d), it is evident that under weather conditions with relatively stable cloud layers, the forecasting discrepancies among the methods—those based solely on TL, those that incorporate TL and utilize raw TF as a part of their input data, and our proposed method which employs TL model while integrating reconstructed TF into the input data—are not markedly pronounced. This is primarily because, under these weather conditions, the variations in the cloud layers are relatively minimal, and the different forecasting accuracy is mainly influenced by the ability of capture subtle cloud variations. Under such weather conditions, there is no obviously different in other factors, indicating that the method proposed in this study possesses a high capability to capture subtle cloud variations.

3.5.3. Forecasting performance with different volume of training data

Several factors can contribute to data loss in PV power stations. The primary reasons for data loss in regional PV power stations include hardware failure [50], where devices like sensors and inverters malfunction; communication issues that result in data transmission interruptions or network outages; and power interruptions which can halt data collection and transmission devices. The primary objective in regional PV power station forecasting is addressing the variance in data volume across different PV stations. Thus, it is imperative to evaluate the performance across PV stations with differing data volumes [51]. The experiment assumes that the dataset of source domain PV station is sufficient, and the data volume of the source domain is set to 1. Tests are conducted when the data volume of the target domain PV station varies as a proportion of the source domain PV station's data volume. In

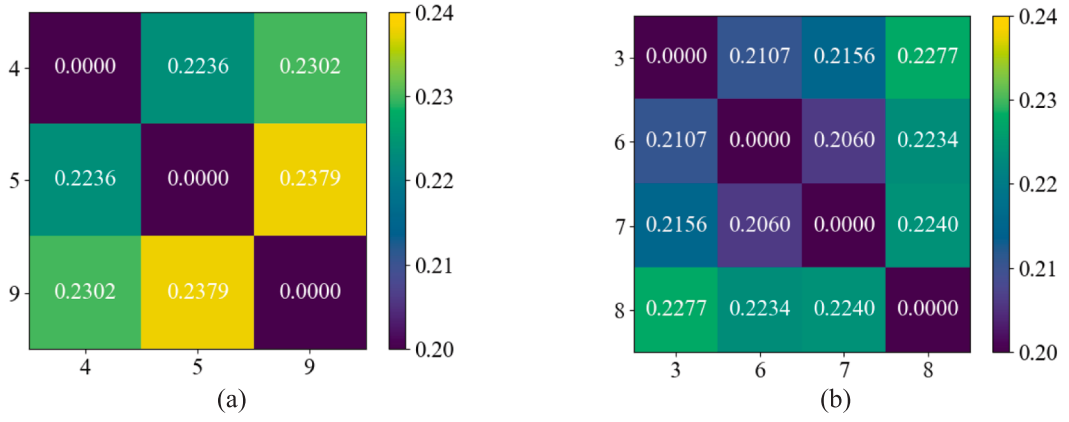


Fig. 16. Heatmap of MMD values between different PV stations. (a) cluster A. (b) cluster B.

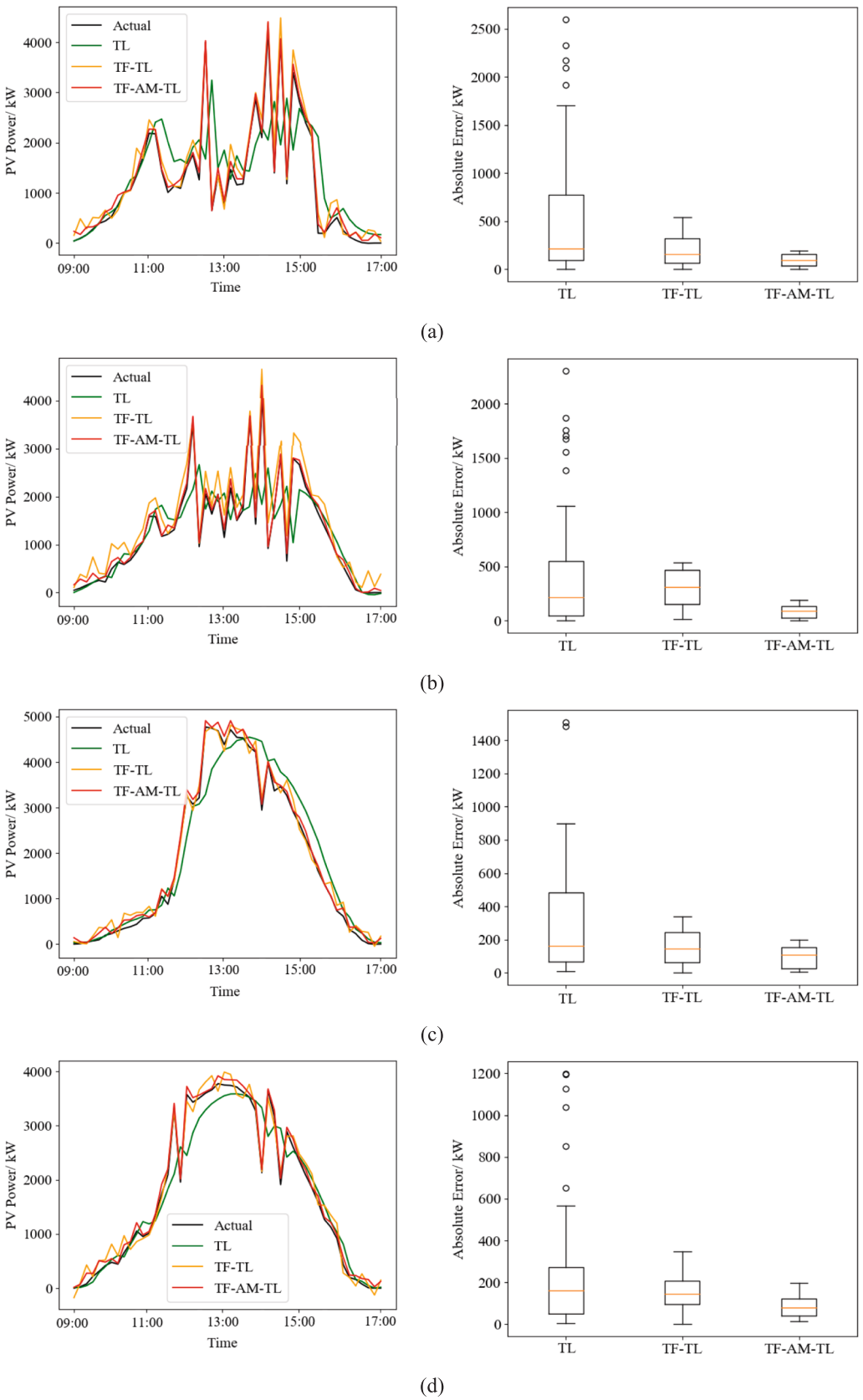


Fig. 17. Forecasting results under different weather condition. (a) No. 4 source PV station under cloudy weather. (b) No. 9 target PV station under cloudy weather. (c) No. 4 source PV station under sunny weather. (d) No. 9 target PV station under sunny weather.

comparative experiments, individual modeling methods are employed to train and test the target domain PV station. Additionally, the method TL and TF-TL and the proposed TF-AM-TL, are comparatively tested. The test results are presented in Table 6.

From the experimental results, it can be observed that the predictive method using the TL model, compared to the standalone modeling predictive method, exhibits superior overall forecasting accuracy, especially when the data volume is at 20 % and 40 %. However, when the data volume exceeds 60 %, the improvement in performance is not significant. This indicates that in the power forecasting of regional PV stations, the TL model demonstrates relatively better performance when there is a lack of data.

To further evaluate the performance of the proposed method in regional PV power forecasting, we compare it with state-of-the-art methods referenced in [20,21], and [22]. It is important to note that all methods were conducted using the transfer learning strategy. The average forecasting results under different volumes of training data are summarized in Fig. 18.

It is evident that the forecasting performance of the method referenced in [20] is the poorest. The global features extracted from the central PV sites do not adequately represent all regional PV sites, and this method fails to capture the unique characteristics of each site. In contrast, the method referenced in [22] demonstrates better performance than that of reference [20] because the PerfCNN module effectively captures weather features across the regional PV sites. By comparison, the proposed method achieves the best performance in regional PV power forecasting among these sites. The TF provides comprehensive information across diverse PV sites, and the proposed transfer learning strategy significantly enhances the capability of feature extraction, taking into account the variations of diverse PV sites and dynamic weather conditions.

3.5.4. Forecasting performances based on data of different seasons

This section examines the forecasting performance of the proposed model for regional PV power based on data from different seasons. Assessing forecasting performance across varying seasons is crucial, as weather conditions and sunshine duration fluctuate significantly throughout the year [52]. To assess the model's generation capabilities and reliability under these diverse conditions, a complete dataset spanning one year was collected to support the experiments.

The dataset for this section was collected from January 14, 2021, to October 24, 2022. Each season was trained using 7 months of data for the comparison models, with the last month used for testing. For example, for the summer season, the data ranged from January 14, 2021, to August 15, 2021. Specifically, the training data for the comparison models spanned from January 14, 2021, to July 14, 2021, while the test data covered the period from July 15, 2021, to August 15, 2021. Similarly, for the autumn season, the training data ranged from March 15, 2021, to November 1, 2021. We present the fluctuation patterns of photovoltaic (PV) power for randomly selected samples across different seasons in Fig. 19 (a)-(d).

The comparison models in this section are all trained using data directly obtained from satellite cloud images. In contrast, the proposed model utilizes time TF constructed by the algorithm developed in this study, along with normal data such as historical PV power, temperature, sunshine duration, and irradiance. The references for the comparison models and their corresponding forecast results are presented in Table 7.

It can be observed that the forecasting performance of the proposed method is superior to that of the other models, maintaining stable performance across diverse seasons. The superiority of the proposed method can be attributed to the time TF constructed in this research and the CNN-INF forecasting model. The TF provide more comprehensive information about cloud motion compared to data obtained directly from satellite cloud images, while the hybrid CNN-INF model effectively extracts spatial and temporal correlations from the fused input feature data.

3.5.5. Comparison of computation efficiency

To compare the computational time, these experiments are conducted on the same platform. The software in the experimental platform of this research is built based on TensorFlow and Pytorch in the python language, and the hardware platform is built with Intel core i7-11700 (CPU) and NVIDIA GeForce GTX 1660 Ti (GPU).

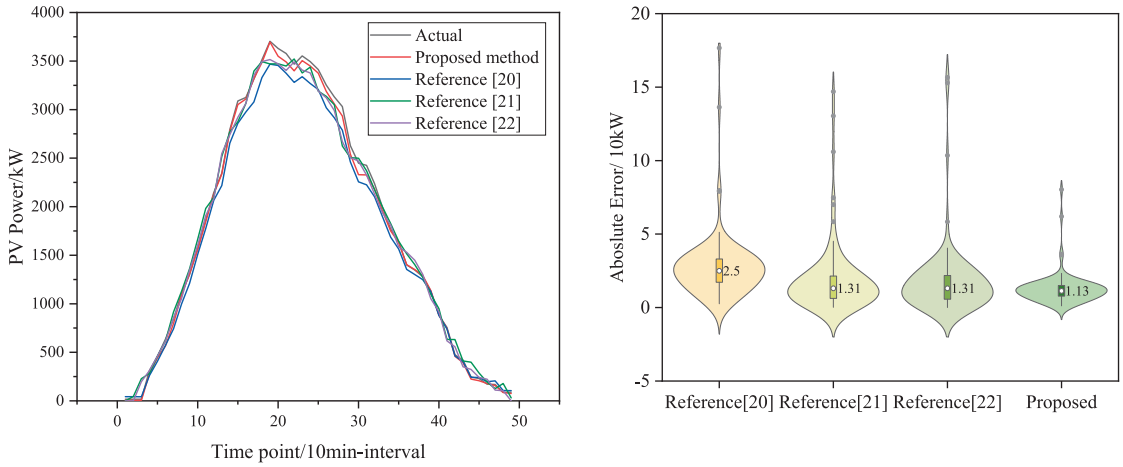
Due to the dispersed relative positions of regional PV stations and their large numbers, enhancing the computational efficiency of their power forecasting is a crucial step towards improving the energy management of regional PV stations. Therefore, testing the computational time for an equivalent number of PV stations is essential. In the experimental section, the No. 3, No. 6, No. 7, and No. 8 PV stations in cluster A are used as examples. We separately tallied the time taken using the standalone modeling method and the method presented in this paper to achieve the same forecasting accuracy. The training loss curve of only target and TL strategy is shown in Fig. 20. And the average computational times are presented in Table 8.

It is obvious that the TL approach significantly reduced the computational time, the computational time approximately improved 54.4 % compared with the standalone modelling method. Illustrated in Fig. 17, the TL model, utilizing pre-trained parameters, achieves lower loss in less than 30 iterations, markedly faster than the 80 iterations for the target-specific model. This rapid convergence highlights TL's effectiveness in optimizing PV forecasting accuracy and accelerating model development.

From the table 8, it can be observed that when computing for a single PV station, the computational time for both standalone modeling and TL algorithms is roughly similar. However, when the number of PV stations computed is more than one, the computational efficiency of the TL model is significantly higher than that of the standalone modeling method.

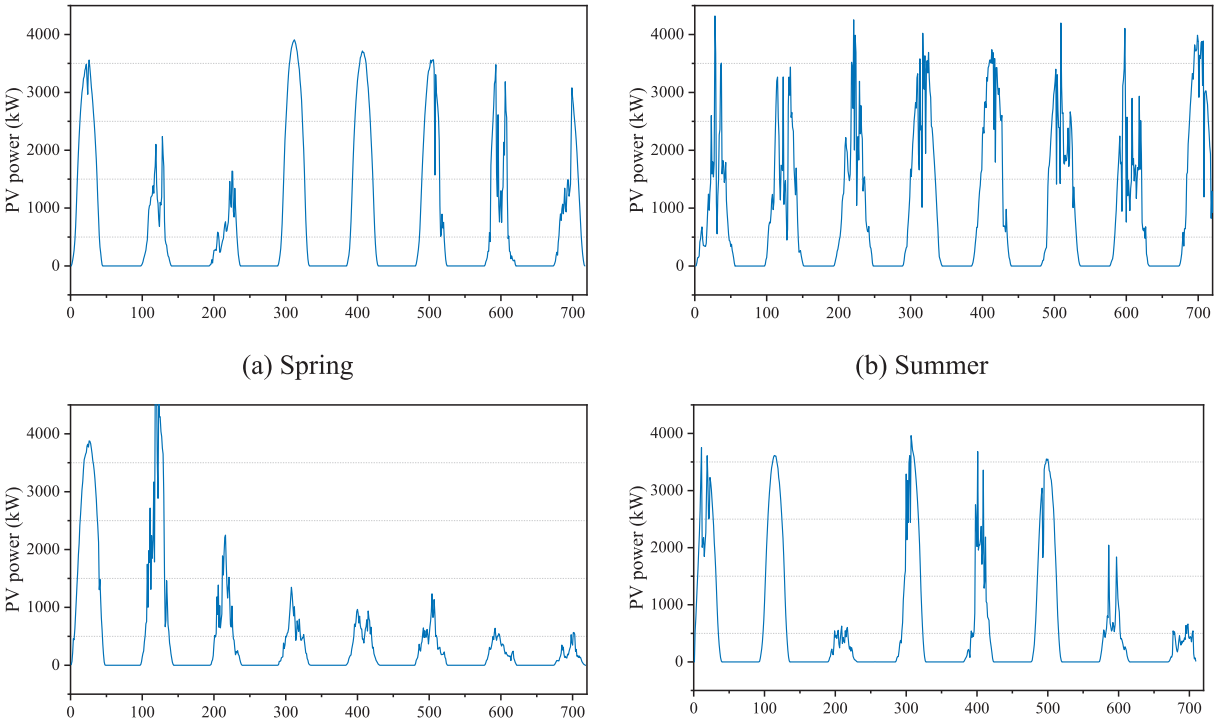
Table 6
Forecasting results of regional PV stations at different data proportions.

Weather conditions	Data proportion (%)	Metrics	Standalone modeling	TL	TF-TL	TF-AM-TL
Cloudy	20	R^2	0.402	0.505	0.552	0.505
		MAE	0.350	0.101	0.058	0.045
		RMSE	0.672	0.230	0.127	0.103
	40	R^2	0.421	0.521	0.694	0.741
		MAE	0.135	0.097	0.051	0.033
		RMSE	0.341	0.191	0.116	0.058
	60	R^2	0.543	0.653	0.782	0.855
		MAE	0.062	0.084	0.055	0.027
		RMSE	0.132	0.185	0.103	0.059
	80	R^2	0.681	0.854	0.903	0.941
		MAE	0.033	0.081	0.053	0.022
		RMSE	0.061	0.153	0.102	0.057
	100	R^2	0.755	0.894	0.941	0.952
		MAE	0.023	0.081	0.052	0.021
		RMSE	0.057	0.152	0.100	0.055
Sunny	20	R^2	0.241	0.425	0.501	0.513
		MAE	0.332	0.050	0.044	0.036
		RMSE	0.576	0.101	0.089	0.071
	40	R^2	0.401	0.653	0.701	0.774
		MAE	0.101	0.032	0.028	0.021
		RMSE	0.235	0.061	0.056	0.045
	60	R^2	0.668	0.785	0.811	0.841
		MAE	0.049	0.027	0.025	0.018
		RMSE	0.093	0.053	0.043	0.033
	80	R^2	0.784	0.821	0.892	0.950
		MAE	0.025	0.022	0.017	0.014
		RMSE	0.047	0.046	0.039	0.031
	100	R^2	0.825	0.848	0.903	0.951
		MAE	0.011	0.022	0.012	0.010
		RMSE	0.023	0.054	0.031	0.024



(a) The 8-hours ahead forecasting results (b) The absolute error distribution

Fig. 18. Comparison of forecasting results with sta te-of-the-art regional PV power forecasting methods.



(a) Spring (b) Summer

(c) Autumn (d) Winter

Fig. 19. Regional PV power data randomly selected from different seasons.

4. Conclusion

In this work, an innovative framework based on TL and satellite cloud imagery has been proposed to enhance the accuracy and computational efficiency of regional PV power forecasting. In the proposed framework, considering of the impacts of cloud motion, an algorithmic architecture composed of GLCM and RF is established for extracting TF from satellite cloud images and selecting the TF with highest correlation to irradiance. And an AM based on LSTM is utilized reconstruct the significant TF. The reconstructed TF obtained from these modules of this framework can provide an extra data for the forecasting model and improve the forecasting accuracy significantly.

Besides, the forecasting module based on TL combined with CNN and INF can address the problem of computational burdens stemming from the vast number and dispersed nature of regional PV installations. Experiments based on single and regional PV power stations are conducted to validate the effectiveness of the proposed framework. The experimental results in the single PV station show that the proposed method has superior performance comparing with the baselines in different weather conditions, and the maximum improvement is approximately 8.7 % in term of RMSE. Furthermore, the results of experiments conducted in the regional PV stations demonstrated that the proposed forecasting model based on TL can improve the computational efficiency obviously by fewer iterations and less volume of training data, and the maximum improvement is approximately 54.4 % in term of computational time and the volume of training data is reduced to 60 %.

Table 7
Forecasting performance for regional PV power in different seasons.

Season	Forecasting model	R^2	MAE	RMSE
Spring	STUNet	0.825	0.089	0.134
	PCPOW	0.857	0.035	0.074
	SVM-INF	0.904	0.021	0.034
	Proposed	0.922	0.008	0.017
Summer	STUNet	0.857	0.037	0.078
	PCPOW	0.912	0.015	0.033
	SVM-INF	0.924	0.012	0.027
	Proposed	0.961	0.013	0.019
Autumn	STUNet	0.922	0.019	0.037
	PCPOW	0.907	0.185	0.031
	SVM-INF	0.912	0.017	0.025
	Proposed	0.962	0.009	0.017
Winter	STUNet	0.894	0.227	0.041
	PCPOW	0.901	0.185	0.035
	SVM-INF	0.914	0.015	0.032
	Proposed	0.954	0.009	0.021

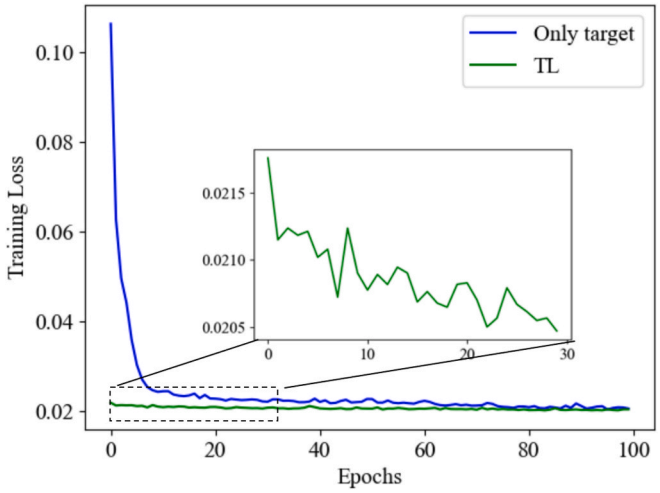


Fig. 20. The training loss curves of Only target and TL models.

Table 8
Computational time comparison between TL and standalone modeling methods.

Computed PV stations No.	Computational time (s) Standalone modeling	Transfer strategy
3	778	786
3, 6, 7, and 8	2265	1032

The superior performance of the proposed forecasting framework is mainly reflected in its excellent accuracy under cloudy weather and its high computational efficiency when the number of the PV power stations is large. Where, the superior forecasting accuracy under cloudy weather condition benefits from algorithmic architecture of TF extraction and reconstruction, as result of these TF can helpful to capture the motion of the cloud that can influence the irradiance above the PV stations directly. And the hybrid deep learning model blending CNN and Informer can not only capture the feature of the reconstructed input data, but also have superior capability for time series forecasting. And the outperformance of computational efficiency under the regional PV stations benefits from the TL-based forecasting model. Because the parameters of the feature extraction module in the source forecasting model can be transferred to target model and the parameters of the fully connection layer can be finetuned by retained in the target dataset. While the proposed method has demonstrated superior performance in regional PV power forecasting in terms of accuracy and computational efficiency under various weather conditions, the forecasting accuracy does decline for horizons exceeding 6 h. Future work will focus on exploring and improving the forecasting performance for longer horizons.

CRediT authorship contribution statement

Yang Xie: Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Jianyong Zheng:** Writing – review & editing, Supervision, Conceptualization. **Mei Fei:** Writing – review & editing, Data curation. **Gareth Taylor:** Writing – review & editing, Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the Science and Technology Program (Key R&D Program) of Guangzhou, China, Grant Number: 2023B01J0004.

Data availability

I have share a link to my data

References

[1] Hussain A, Arif SM, Aslam M. Emerging renewable and sustainable energy technologies: State of the art. Renew Sustain Energy Rev 2017;71:12–28.
[2] Allouhi A, Rehman S, Buker MS, Said Z. Recent technical approaches for improving energy efficiency and sustainability of PV and PV-T systems: a comprehensive review. Sustainable Energy Technol Assess 2023;56:103026.
[3] Lindahl J, Lingfors D, Elmqvist Å, Mignon I. Economic analysis of the early market of centralized photovoltaic parks in Sweden. Renew Energy 2022;185:1192–208.
[4] Wang R, Hasanefendic S, Von Hauff E, Bossink B. The cost of photovoltaics: Re-evaluating grid parity for PV systems in China. Renew Energy 2022;194:469–81.
[5] Li P, Zhou K, Lu X, Yang S. A hybrid deep learning model for short-term PV power forecasting. Appl Energy 2020;259:114216.
[6] Wang F, Lu X, Mei S, Su Y, Zhen Z, Zou Z, et al. A satellite image data based ultra-short-term solar PV power forecasting method considering cloud information from neighboring plant. Energy 2022;238:121946.
[7] Quiñones JJ, Pineda LR, Ostanek J, Castillo L. Towards smart energy management for community microgrids: Leveraging deep learning in probabilistic forecasting of renewable energy sources. Energ Conver Manage 2023;293:117440.
[8] Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew Sustain Energy Rev 2020;124:109792.
[9] Huang C, Shi H, Yang D, Gao L, Zhang P, Fu D, et al. Retrieval of sub-kilometer resolution solar irradiance from Fengyun-4A satellite using a region-adapted Heliosat-2 method. Sol Energy 2023;264:112038.
[10] Kong W, Jia Y, Dong ZY, Meng K, Chai S. Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting. Appl Energy 2020;280:115875.
[11] Zhen Z, Pang S, Wang F, Li K, Li Z, Ren H, et al. Pattern classification and PSO optimal weights based sky images cloud motion speed calculation method for solar PV power forecasting. IEEE Trans Ind Appl 2019;55:3331–42.
[12] Jiang J, Lv Q, Gao X. The ultra-short-term forecasting of global horizonal irradiance based on total sky images. Remote Sens (Basel) 2020;12:3671.
[13] Hu K, Wang L, Li W, Cao S, Shen Y. Forecasting of solar radiation in photovoltaic power station based on ground-based cloud images and BP neural network. IET Generation, Transmission & Distribution 2022;16:333–50.
[14] Ren S, Hu W, Bradbury K, Harrison-Atlas D, Valeri LM, Murray B, et al. Automated extraction of energy systems information from remotely sensed data: a review and analysis. Appl Energy 2022;326:119876.
[15] Yang L, Gao X, Li Z, Jia D, Jiang J. Nowcasting of surface solar irradiance using FengYun-4 satellite observations over China. Remote Sens (Basel) 2019;11:1984.
[16] Wang F, Li J, Zhen Z, Wang C, Ren H, Ma H, et al. Cloud feature extraction and fluctuation pattern recognition based ultrashort-term regional PV power forecasting. IEEE Trans Ind Appl 2022;58:6752–67.
[17] Wang K, Shan S, Dou W, et al. A cross-modal deep learning method for enhancing photovoltaic power forecasting with satellite imagery and time series data. Energ Conver Manage 2025;323:119218.

[18] Zhen Z, Pang S, Wang F, et al. Pattern classification and PSO optimal weights based sky images cloud motion speed calculation method for solar PV power forecasting. IEEE Trans Ind Appl 2019;55(4):3331–42.
[19] Nie B, Lu Z, Han J, et al. Investigation on Ground-based Cloud image Classification and its Application in Photovoltaic Power Forecasting. IEEE Trans Instrum Meas 2025.
[20] Shi J, Chen Y, Cheng X, et al. Four-stage space-time hybrid model for distributed photovoltaic power forecasting. IEEE Trans Ind Appl 2022;59(1):1129–38.
[21] Li J, Rao C, Gao M, et al. Efficient calculation of distributed photovoltaic power generation power prediction via deep learning. Renew Energy 2025;246:122901.
[22] Lai W, Zhen Z, Wang F, et al. Sub-region division based short-term regional distributed PV power forecasting method considering spatio-temporal correlations. Energy 2024;288:129716.
[23] Li X, Ma L, Chen P, Xu H, Xing Q, Yan J, et al. Probabilistic solar irradiance forecasting based on XGBoost. Energy Rep 2022;8:1087–95.
[24] Wang Y, Liao W, Chang Y. Gated recurrent unit network-based short-term photovoltaic forecasting. Energies 2018;11:2163.
[25] Akhter MN, Mekhilef S, Mokhlis H, Mohamed SaN. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET Renew Power Gener 2019;13:1009–23.
[26] Ghimire S, Deo RC, Raj N, Mi J. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 2019;253:113541.
[27] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: a review. Energ Conver Manage 2018;156:459–97.
[28] Tapakis R, Charalambides AG. Equipment and methodologies for cloud detection and classification: a review. Sol Energy 2013;95:392–430.
[29] Si Z, Yang M, Yu Y, Ding T. Photovoltaic power forecast based on satellite images considering effects of solar position. Appl Energy 2021;302:117514.
[30] Xu X, Zhang L, Yang J, Cao C, Tan Z, Luo M. Object detection based on fusion of sparse point cloud and image information. IEEE Trans Instrum Meas 2021;70:1–12.
[31] Shabat AM, Tapamo J. A comparative study of the use of local directional pattern for texture-based informal settlement classification. J Appl Res Technol 2017;15:250–8.
[32] Xu Y, Xie Y, Wang X, Shen X, Song M, Hang W. Study on a novel defrost control method based on the surface texture of evaporator image with gray-level cooccurrence matrix, new characterization parameter combination and machine learning. Energ Buildings 2023;292:113173.
[33] Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. IEEE Trans Syst Man Cybern 1973:610–21.
[34] Chen Z, Han F, Wu L, Yu J, Cheng S, Lin P, et al. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. Energ Conver Manage 2018;178:250–64.
[35] Jin CH, Pok G, Lee Y, Park H, Kim KD, Yun U, et al. A SOM clustering pattern sequence-based next symbol prediction method for day-ahead direct electricity load and price forecasting. Energ Conver Manage 2015;90:84–92.
[36] Aponte O, McConky KT. Forecasting an electricity demand threshold to proactively trigger cost saving demand response actions. Energ Buildings 2022;268:112221.
[37] Wang F, Zhen Z, Liu C, Mi Z, Hodge B, Shafie-khah M, et al. Image phase shift invariance based cloud motion displacement vector calculation method for ultra-short-term solar PV power forecasting. Energ Conver Manage 2018;157:123–35.
[38] Ramahatana F, La Salle JLG, Lauret P, David M. A more efficient microgrid operation through the integration of probabilistic solar forecasts. Sustainable Energy Grids Networks 2022;31:100783.
[39] Shimomura M, Keeley AR, Matsumoto K, Tanaka K, Managi S. Beyond the merit order effect: Impact of the rapid expansion of renewable energy on electricity market price. Renew Sustain Energy Rev 2024;189:114037.
[40] Jiang H, Lu N, Huang G, Yao L, Qin J, Liu H. Spatial scale effects on retrieval accuracy of surface solar radiation using satellite data. Appl Energy 2020;270:115178.
[41] Shan S, Li C, Ding Z, Wang Y, Zhang K, Wei H. Ensemble learning based multi-modal intra-hour irradiance forecasting. Energ Conver Manage 2022;270:116206.
[42] Shan S, Ding Z, Zhang K, Wei H, Li C, Zhao Q. ACGL-TR: a deep learning model for spatio-temporal short-term irradiance forecast. Energ Conver Manage 2023;284:116970.
[43] Wang K, Qi X, Liu H. Photovoltaic power forecasting based LSTM-Convolutional Network. Energy 2019;189:116225.
[44] Guermoui M, Bouchouicha K, Bailek N, Boland JW. Forecasting intra-hour variance of photovoltaic power using a new integrated model. Energ Conver Manage 2021;245:114569.
[45] Zang H, Xu R, Cheng L, Ding T, Liu L, Wei Z, et al. Residential load forecasting based on LSTM fusing self-attention mechanism with pooling. Energy 2021;229:120682.
[46] Visser L, AlSkaif T, van Sark W. Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. Renew Energy 2022;183:267–82.
[47] Ilias L, Sarmas E, Marinakis V, Askounis D, Doukas H. Unsupervised domain adaptation methods for photovoltaic power forecasting. Appl Soft Comput 2023;149:110979.
[48] Tang Y, Yang K, Zhang S, Zhang Z. Photovoltaic power forecasting: a hybrid deep learning model incorporating transfer learning strategy. Renew Sustain Energy Rev 2022;162:112473.
[49] Shi H, Yang D, Wang W, Fu D, Gao L, Zhang J, et al. First estimation of high-resolution solar photovoltaic resource maps over China with Fengyun-4A satellite and machine learning. Renew Sustain Energy Rev 2023;184:113549.
[50] Qin J, Jiang H, Lu N, Yao L, Zhou C. Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning. Renew Sustain Energy Rev 2022;167:112680.
[51] Rohani A, Tak i M, Abdollahpour M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). Renewable Energy 2018; 115:411-22.
[52] Ghorvei M, Kavianpour M, Beheshti MT, Ramezani A. Spatial graph convolutional neural network via structured subdomain adaptation and domain adversarial learning for bearing fault diagnosis. Neurocomputing 2023;517:44–61.

Yang Xie received his B.S. degree in Electrical Engineering and Automation from the Shenyang Institute of Engineering, Shenyang, China, in 2015, his M.S. degree in Electrical Engineering from Shenyang University of Technology, Shenyang, China, in 2018, and his Ph.D. degree in Cybersecurity from Southeast University, Nanjing, China. He is currently a Lecturer at Guangdong Polytechnic Normal University. His research interests include artificial intelligence applications in power systems, renewable energy output forecasting, and frequency security in modern power systems.

Jianyong Zheng is a professor at the School of Electrical Engineering and the School of Cyber Science and Engineering at Southeast University. He is the leader of the Department of Electrical Engineering.

Fei Mei is a lecturer at Hohai University. His research focuses on artificial intelligence and big data mining technologies for power systems, electrical equipment fault diagnosis, and power quality.

Gareth Taylor joined Brunel University in 2000, became a lecturer in 2003, and directed the MSc Sustainable Electrical Power program from 2006. He led the Brunel University Research Centre BIPS from 2010 and was named Professor in Power Systems in 2012. He headed the Department of Electronic and Electrical Engineering from 2019 to 2023 and has held significant roles in IEEE, IET, and CIGRE. Taylor also served as a Visiting Professor at Imperial College London from 2023 to 2026.