Energy Engineering

Tech Science Press

DOI: 10.32604/ee.2025.059533

ARTICLE

Short-Term Photovoltaic Power Prediction Based on Multi-Stage Temporal Feature Learning

Qiang Wang[1], Hao Cheng[2], Wenrui Zhang[2,*], Guangxi Li[3], Fan Xu[2], Dianhao Chen[4] and Haixiang Zang[4]

[1]Jiangsu Qitian Power Construction Group Co., Ltd., Lianyungang, 222000, China

[2]Lianyungang Zhiyuan Electricity Design Co., Ltd., Lianyungang, 222000, China

[3]Lianyungang Power Supply Company, State Grid Jiangsu Electric Power Co., Ltd., Lianyungang, 222000, China

[4]School of Electrical and Power Engineering, Hohai University, Nanjing, 210098, China

[*]Corresponding Author: Wenrui Zhang. Email: wenruizhang_lyg@163.com

Received: 10 October 2024 Accepted: 19 December 2024 Published: 31 January 2025

ABSTRACT

Harnessing solar power is essential for addressing the dual challenges of global warming and the depletion of traditional energy sources. However, the fluctuations and intermittency of photovoltaic (PV) power pose challenges for its extensive incorporation into power grids. Thus, enhancing the precision of PV power prediction is particularly important. Although existing studies have made progress in short-term prediction, issues persist, particularly in the underutilization of temporal features and the neglect of correlations between satellite cloud images and PV power data. These factors hinder improvements in PV power prediction performance. To overcome these challenges, this paper proposes a novel PV power prediction method based on multi-stage temporal feature learning. First, the improved LSTM and SA-ConvLSTM are employed to extract the temporal feature of PV power and the spatial-temporal feature of satellite cloud images, respectively. Subsequently, a novel hybrid attention mechanism is proposed to identify the interplay between the two modalities, enhancing the capacity to focus on the most relevant features. Finally, the Transformer model is applied to further capture the short-term temporal patterns and long-term dependencies within multi-modal feature information. The paper also compares the proposed method with various competitive methods. The experimental results demonstrate that the proposed method outperforms the competitive methods in terms of accuracy and reliability in short-term PV power prediction.

KEYWORDS

Photovoltaic power prediction; satellite cloud image; LSTM-Transformer; attention mechanism

1 Introduction

1.1 Background

Progress in the growth of photovoltaic (PV) energy production is vital for easing the present-day energy shortage and mitigating environmental impacts [1]. However, integrating large-scale PV systems into the grid introduces potential challenges to the security and stability of the power system. This is primarily due to the inherent fluctuations and intermittency of PV generation, which can lead to issues such as frequency deviations and harmonic distortions [2]. Hence, it is crucial to develop an accurate prediction method to tackle these issues and guarantee the stability and reliability of the electrical power system.

1.2 Related Works

The variability and unpredictability of photovoltaic (PV) energy output are largely due to changing weather patterns. Consequently, traditional forecasting techniques frequently include meteorological data as key variables in their models. Literature [3] utilized multi-scale permutation entropy to characterize the power state under diverse weather conditions, thereby reducing the forecasting sequence's susceptibility to meteorological influences. Literature [4] introduced a dual-similarity day selection model that categorizes weather into three distinct classes, enhancing the data's quality and reliability. However, some studies have proved that the movement of clouds is the main reason for fluctuations in PV power generation [5–7]. Data regarding cloud conditions can be derived from both terrestrial and satellite imagery, which is particularly useful for extensive short-term or medium-term weather predictions, as well as for hourly forecasts [8]. These images provide a wealth of details, such as the spatial layout, luminosity, and shape [9]. Satellite cloud images, with their broader space-time field of view compared to ground-based sky images, are particularly well-suited for long-horizon prediction tasks [10]. Consequently, incorporating satellite cloud imagery into the study of short-term PV power prediction is of significant importance.

Many studies have conducted in-depth research on image-driven photovoltaic power methods. Convolutional Neural Networks (CNNs), frequently utilized as image feature extractors within the domain of computer vision, have garnered widespread application in cloud studies. Literature [11] applied various deep fully CNNs as feature extractors, demonstrating that the different depths of the fully CNNs can enhance the predictive accuracy and stability of the model. Literature [12] used the CNNs with residual structure to extract the image features, which preserves as much valid information as possible for the prediction task. Furthermore, several studies have identified that the spatial-temporal features of historical multiple images can further enhance prediction performance. Literature [7] introduced an auto-encoder (CAE) that employs three-dimensional CNNs (3DCNN) to address the shortcomings of conventional forecasting models, such as the constraints on the length of input image sequences and the limitations of linear image extrapolation. Literature [13] presented a 3D convolutional long-short-term memory (ConvLSTM)-CNN hybrid framework, which extracts spatial-temporal features from multiple images across various color spaces to improve forecasting performances. Nevertheless, due to the structural limitations of models such as 3DCNN and ConvLSTM, spatial-temporal feature information extraction may not be fully adequate. Therefore, it is necessary to further refine the architecture of such models to achieve performance improvements.

Moreover, PV power has a significant temporal correlation, so many studies focus on temporal feature extraction [14–16]. Literature [17] proposed an improved gated recurrent unit containing RepeatVector and TimeDistributed layer, which significantly boosts the precision of PV power forecasting. Literature [18] utilized the Coati optimization algorithm to update the hyperparameters of the long-short-term memory (LSTM), which leads to a distinctive improvement of 16%∼36% compared to other baseline methods. Literature [19] has demonstrated that the fusion of LSTM and temporal convolutional neural network (TCN) exhibits enhanced capability in capturing temporal features in comparison to using LSTM and TCN separately. Literature [20] proposed a causal convolutional Transformer, which enhances the extraction of both global and local features by integrating a linear embedding module and a causal convolutional module. The aforementioned methods fully demonstrate the importance of temporal features. A reasonable model combination is a feasible solution to further improve the performance. The combination of LSTM and Transformer has proven its effectiveness [21,22]. Literature [23] employed pre-trained LSTM models to acquire weather forecasting results, which were utilized as auxiliary inputs for the Transformer to mitigate forecast result uncertainty. In literature [24], LSTM and informer were used in parallel to extract local and global temporal features, which significantly boosts the accuracy and effectiveness of the sequence modeling task.

1.3 Research Gaps and Contributions

While the LSTM-Transformer model has demonstrated effectiveness in enhancing the precision of PV power prediction, there remain unresolved issues that warrant improvement. One such issue pertains to the exclusive utilization of the LSTM-Transformer structure for historical PV power feature engineering. Given the analogous temporal correlations present in historical satellite cloud images, there is potential value in extending the application of the LSTM-Transformer structure to the feature engineering of images. Furthermore, it should be recognized that both LSTM and ConvLSTM models, as previously discussed in the literature [25], are prone to the challenges of vanishing gradients and inadequate capabilities in capturing long-range dependencies. These challenges can significantly impede the training efficacy and overall performance of models incorporating the LSTM-Transformer architecture.

Moreover, the approach to predict utilizing historical PV data and satellite cloud images involves two distinct modal inputs, each with specific feature extractors. This setup presents challenges in effectively capturing the complementary and redundant aspects of multi-modal data. Consequently, the fusion of multi-modal features has garnered significant attention within the machine learning domain [9]. Traditional feature fusion techniques such as concatenation [26] and addition [27] have limitations as they do not fully account for the interaction information between modalities and cannot adjust fusion strategies dynamically. Over the past few years, the attention mechanism has become a key technique for tackling challenges in feature fusion [28,29]. Therefore, there is a growing need for advancing feature fusion methods based on the attention mechanism for enhancing PV power prediction accuracy.

Motivated by the aforementioned gaps, this study proposes a short-term photovoltaic power prediction method based on multi-stage temporal feature learning. The key contributions of this study are outlined as follows:

- As a key module of the prediction method, an improved LSTM model was developed. Within the improved LSTM, a bidirectional dynamic residual mechanism was proposed to effectively enhance the modeling capability of time-series dynamics and alleviate the vanishing gradient problem.
- A novel hybrid attention mechanism was introduced to comprehensively integrate multi-modal features. This mechanism includes a channel-wise self-attention module designed to emphasize crucial information within each modality, complemented by a residual cross-attention module that explores correlations across different modalities.
- An end-to-end deep learning method for short-term PV power prediction was established. The proposed method was sufficiently validated and compared with various competitive methods, which further confirms the proposed method exhibits higher generalization and robustness.

2 Methodology

2.1 Overall Framework

As shown in Fig. 1, the proposed method consists of three components: the feature extraction module, the feature fusion module, and the temporal analysis module. Initially, the feature extraction module employs an improved LSTM and SA-ConvLSTM to extract the temporal feature of PV power and the spatial-temporal feature of satellite cloud images, respectively. Following this, the feature fusion module introduces a novel hybrid attention mechanism that delves into the coupling correlations. Finally, the temporal analysis module based on Transformer architecture is applied to capture short-term temporal patterns and to model long-term dependencies.

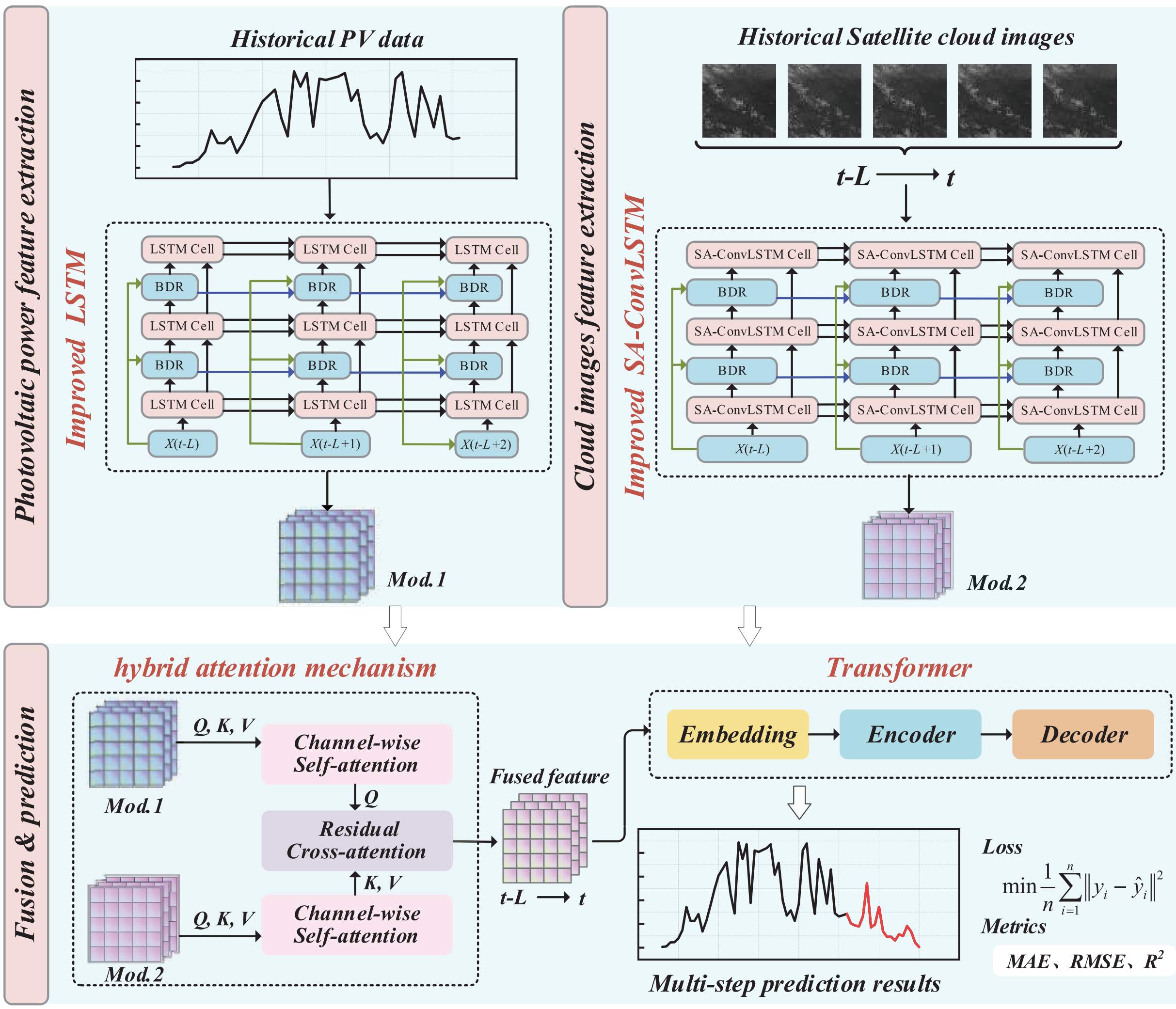

Figure 1: The proposed method

2.2 Feature Extraction Module

Both PV power and satellite cloud images have significant temporal characteristics, so making good use of these characteristics is crucial to improve the accuracy of PV power prediction. This module involves utilizing improved SA-ConvLSTM to capture spatial-temporal features from satellite cloud images and employing improved LSTM to identify temporal patterns from historical PV power. To bolster the effectiveness of SA-ConvLSTM and LSTM, a bidirectional dynamic residual mechanism is introduced in our study.

2.2.1 SA-ConvLSTM

The SA-ConvLSTM is an enhanced version of the ConvLSTM, designed to improve the capacity to capture extensive spatial dependencies. This improvement is realized through the incorporation of a self-attention component within the ConvLSTM framework, as depicted in Fig. 2. In contrast to conventional ConvLSTM networks, SA-ConvLSTM introduces a Self-Attention Memory module (SAM) that refines the self-attention mechanism to retain features with long-term spatial and temporal dependencies, as shown in Fig. 3. Consequently, this modification improves the model's capability to capture spatial-temporal features. The operational framework of SA-ConvLSTM can be articulated as follows:

$$
\begin{aligned}
&\hat{x}_t = SA(x_t), \hat{h}_{t-1} = SA(h_{t-1}) \\
&i_t = \sigma\left(W_{oi} * \hat{x}_t + W_{hi} * \hat{h}_{t-1} + b_i\right) \\
&f_t = \sigma\left(W_{of} * \hat{x}_t + W_{hf} * \hat{h}_{t-1} + b_f\right) \\
&c_t = f_t \otimes c_{t-1} + i_t \otimes \tanh\left(W_{oc} * \hat{x}_t + W_{hc} * \hat{h}_{t-1} + b_c\right) \\
&g_t = \sigma\left(W_{og} * \hat{x}_t + W_{hg} * \hat{h}_{t-1} + b_g\right) \\
&h_t = g_t \otimes \tanh(c_t)
\end{aligned}
\tag{1}
$$

where $\hat{x}_t$, $\hat{h}_t$, c_t represent the input value, output value, and cell state at t-th timestep, respectively; SA is the self-attentional memory module; i_t, f_t, g_t represent the input gates, forget gates, and output gates; W_{xi}, W_{hi}, b_i, W_{xf}, W_{hf}, b_f, W_{xc}, W_{hc}, b_c, W_{xg}, W_{hg}, b_g are trainable parameter; $\sigma(\cdot)$ is activation function, $\otimes$ is Hadamard product.

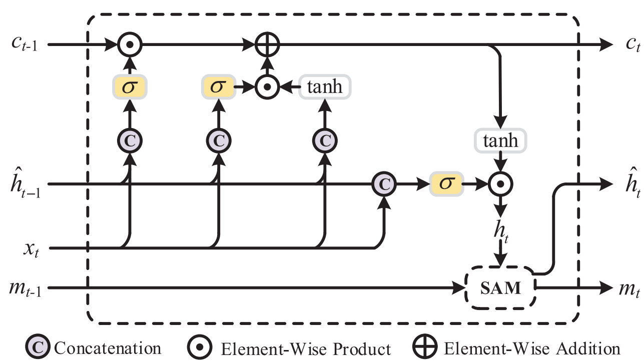

Figure 2: The structure of SA-ConvLSTM

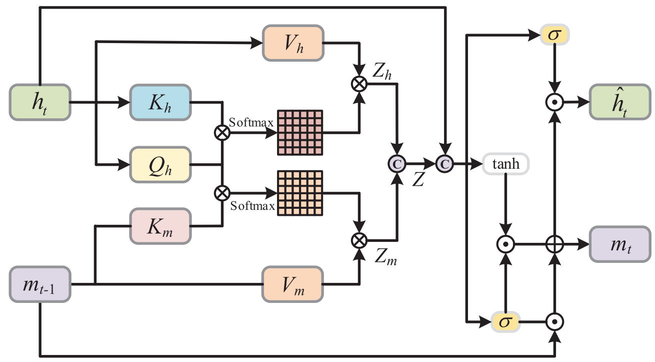

Figure 3: The structure of the self-attention memory module

2.2.2 *Bidirectional Dynamic Residual Mechanism*

In comparison to recurrent neural networks (RNN), while LSTM demonstrates superior performance in tasks involving time series prediction, challenges persist such as the occurrence of vanishing gradients and limitations in long-distance modeling capabilities. Given these problems, this study introduces a novel bidirectional dynamic residual (BDR) mechanism tailored for LSTM and its variants, illustrated in Fig. 4. By incorporating residual structures in both horizontal and vertical orientations, a rapid pathway is established for transmitting crucial feature information, thereby addressing issues related to vanishing and exploding gradients, and augmenting the network's representational and learning capacities. The fundamental component of BDR is depicted in Fig. 5, and its formulation can be articulated as follows:

$$\begin{aligned} v_{t-1}^{1} &= v_{t-1}^{0} + \left[\sigma\left(W_{h1} * h_{t-1}^{1} + W_{v1} * v_{t-1}^{0} + b_{v1}\right) \otimes \tanh\left(W_{h2} * h_{t-1}^{1} + W_{v2} * v_{t-1}^{0} + b_{v2}\right)\right] \\ z_{t-1}^{1} &= z_{t-2}^{1} + \left[\sigma\left(W_{h3} * h_{t-1}^{1} + W_{z1} * z_{t-2}^{1} + b_{z1}\right) \otimes \tanh\left(W_{h4} * h_{t-1}^{1} + W_{z2} * z_{t-2}^{1} + b_{z2}\right)\right] \end{aligned} \tag{2}$$

where $W_{h1}, W_{h2}, W_{h3}, W_{h4}, W_{v1}, W_{v2}, W_{z1}, W_{z2}, b_{v1}, b_{v2}, b_{z1}, b_{z2}$ are trainable parameters. The superscript and subscript of $v_{t-1}^{0}, v_{t-1}^{1}, z_{t-1}^{1}, z_{t-2}^{1}, h_{t-1}^{1}$ is the layer number of LSTM and timestep. Then, the h_t^0 will be converted to:

$$\tilde{h}_{t-1}^{1} = Concat([Concat([h_{t-1}^{1}, z_{t-2}^{1}]), Concat([h_{t-1}^{1}, v_{t-1}^{0}])]) \tag{3}$$

2.3 Feature Fusion Module

This study presents a fusion module designed to integrate multi-modal features, incorporating a channel-wise self-attention mechanism and a residual cross-attention mechanism. The process, illustrated in Fig. 6, involves two sequential steps for each moment. Initially, the channel-wise self-attention mechanism is applied to the single modality feature, followed by the derivation of fusion features through the residual cross-attention mechanism.

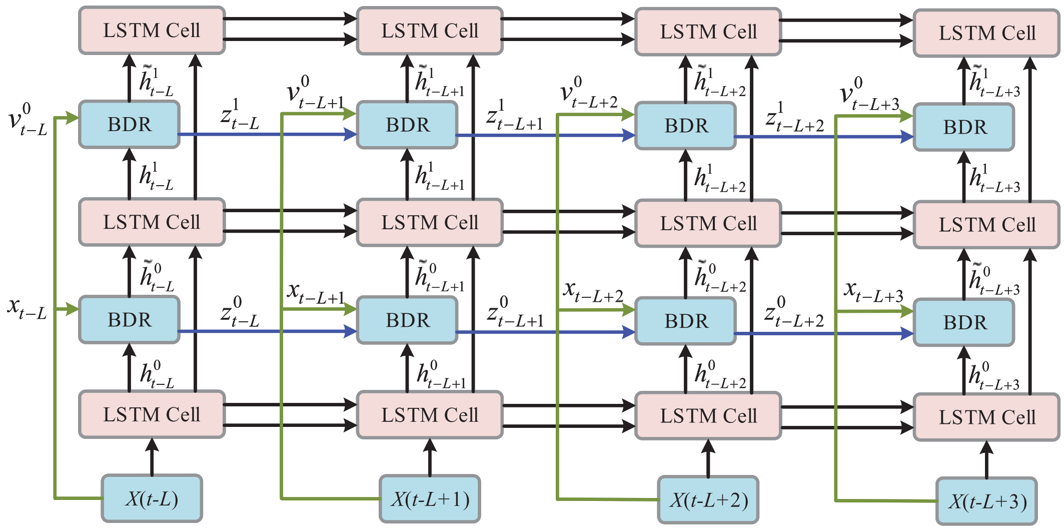

Figure 4: The improved LSTM with the bidirectional dynamic residual mechanism

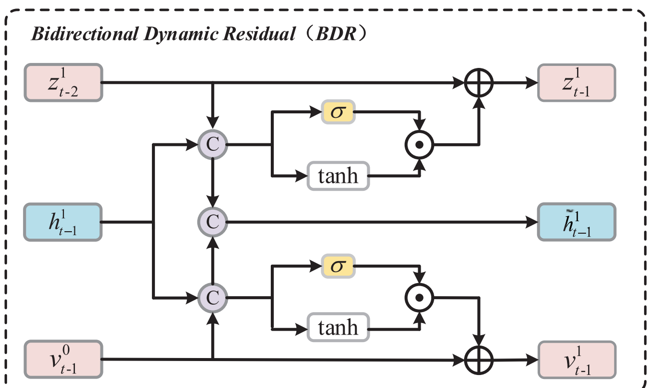

Figure 5: The basic unit of the bidirectional dynamic residual mechanism

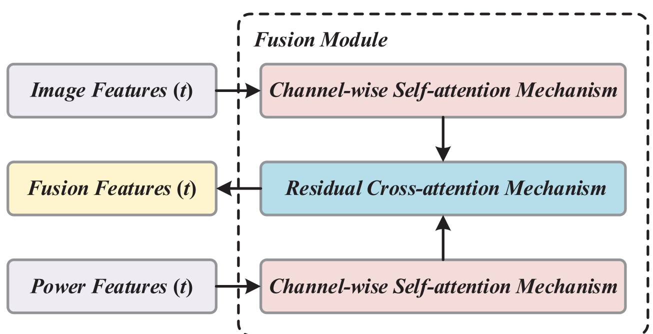

Figure 6: The multi-modal feature fusion module

As shown in Fig. 7, the channel-wise self-attention mechanism is developed to establish correlations among the features present within each channel of each individual modality. Through the allocation of weights to the features, the prediction methods enable a better understanding of the importance of each channel. Specifically, these weights are employed to calibrate and combine the features of each channel, thereby augmenting the comprehension and depiction of the features inherent to the individual modality. The fundamental concept underlying the channel-wise self-attention mechanism can be articulated as follows:

$$S(Q, K, V) = \text{Softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{4}$$

where d_k is the dimension of K. The $\sqrt{d_k}$ is adopted for smoothing the backward gradients. Softmax is an exponential normalization function. Q, K, V can be derived from a single modality feature by non-linear projection.

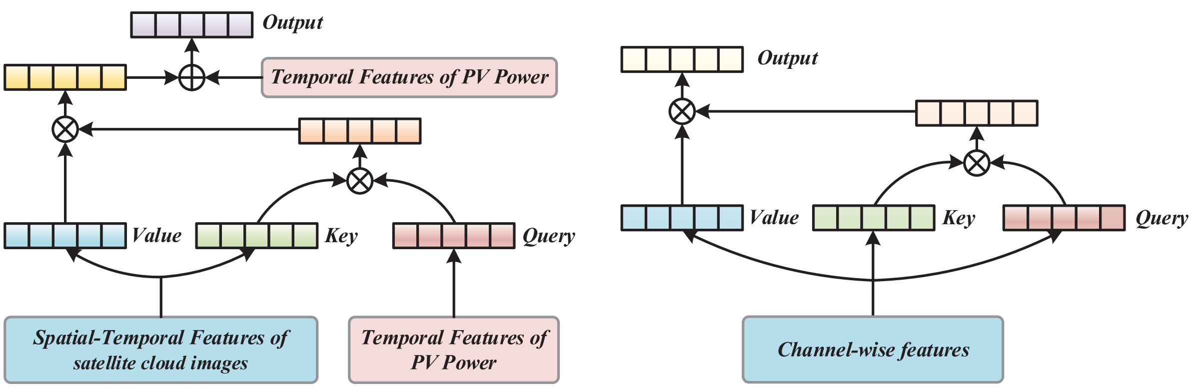

Figure 7: The self-attention and residual cross-attention

The residual cross-attention mechanism, commonly employed in natural language processing and computer vision domains, aids in establishing connections between features from different modalities. This enables the information from one modality to influence the processing of another, thereby improving the model's capability to capture interrelations between the diverse modal features and enhancing the overall representation quality. The residual cross-attention mechanism is formulated as:

$$C(Q, K, V) = \text{Softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V + F \tag{5}$$

where K and V are obtained from image features feature by non-linear projection. Q is obtained from power features feature F by non-linear projection.

2.4 Temporal Analysis Module

As the core of the temporal analysis module, Transformer model incorporates an attention mechanism to overcome the challenge of limited parallelization in RNNs when processing lengthy temporal sequences. As shown in Fig. 8, a Transformer block consists of various components including multi-head attention modules, multi-layer perceptron modules, layer normalization, and residual connections [30]. As a key component of Transformer, the multi-head attention provides benefits such as the parallel focus on various input segments, enhancing attention, and capturing complex dependencies. The multi-head attention mechanism can be represented as:

$$\text{MultiHead}\,(Q, K, V) = Concat\,(head_1, \ldots, head_h)\, W^O, \quad head_i = Attention\left(QW_i^Q, KW_i^K, VW_i^V\right) \tag{6}$$

where $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^V \in \mathbb{R}^{d_{model} \times d_V}$, and $W^O \in \mathbb{R}^{hd_V \times d_{model}}$ are parameters of the linear projections; d_{model} represents the output dimension. In our proposed method, the input of Transformer is the fused feature acquired by the aforementioned feature fusion module. Utilizing the Transformer architecture, the model is capable of capturing both short-term temporal dynamics and long-term dependencies, which are essential factors for accurately predicting PV power output.

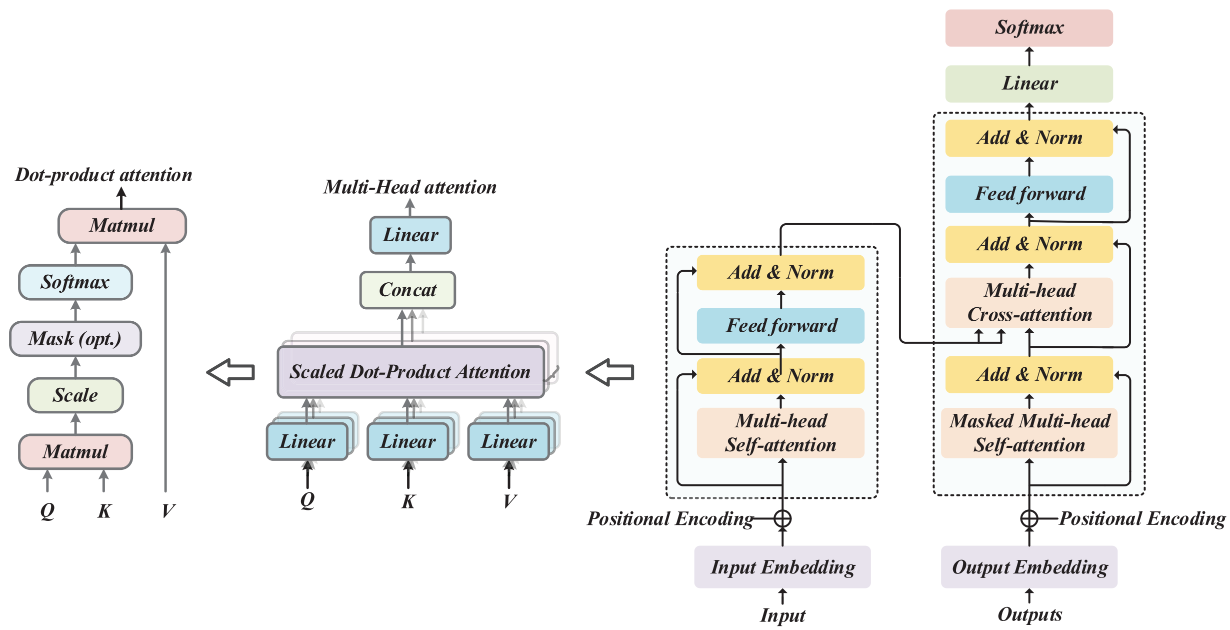

Figure 8: The structure of Transformer

3 Results and Discussions

3.1 Experimental Settings

3.1.1 Data Source

The satellite cloud image data utilized in this study were sourced from the Himawari-8 satellite operated by the Japan Meteorological Agency. The PV power dataset was obtained from the Desert Knowledge Australia (DKA) Solar Centre, with the PV power data representing a capacity of 263.0 kW and collected at 10-min intervals during the period of 2016–2017.

3.1.2 Performance Metrics

To assess the effectiveness of the proposed method, we employ the following evaluation metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). Here are the mathematical representations of these metrics:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{7}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{8}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_i - y_i\right|}{y_i} \times 100\% \tag{9}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{N}\left(\overline{y}_i - y_i\right)^2} \tag{10}$$

where y_i, $\hat{y}_i$, $\overline{y}_i$ represent the actual output, forecast output, and mean value of actual output respectively.

3.1.3 Competitive Methods and Experiment Setting

To ensure the results' objectivity and reliability, this paper conducted ten experiments and averaged the results. Five competitive methods were adopted for comparisons to validate the effectiveness of the proposed method for short-term PV power prediction, as shown in Table 1.

Table 1: Configurations of each method

Methods	Feature extractors for PV power	Feature extractors for satellite image	Feature extractors for fusion feature
S1	Transformer	SA-ConvLSTM	Transformer
*S*2	LSTM	ConvLSTM	Transformer
*S*3	LSTM	SA-ConvLSTM	LSTM
*S*4	LSTM	SA-ConvLSTM	Transformer
*S*5	Improved LSTM	Improved SA-ConvLSTM	LSTM
Proposed	Improved LSTM	Improved SA-ConvLSTM	Transformer

All these methods were developed using Tensorflow and Keras. Parameter optimization was conducted using the Adam optimizer, with the mean square error (MSE) function employed as the loss metric. Training for each method spanned 100 epochs with a batch size of 512. To avoid overfitting, an early stopping strategy with a patience parameter set to 5 epochs was implemented. Additionally, the starting learning rate was set to 0.001.

3.2 Performance Evaluation

3.2.1 Comparisons of the Proposed and Competitive Methods

Based on historical PV power data and satellite cloud images, this paper predicts the photovoltaic power in the next hour, and the prediction results are shown in Table 2. The R^2 of the proposed method is 0.935, and RMSE and MAE are 2.4%∼11.9% and 8.0%∼34.7% higher than competitive methods, respectively. *S*2 exhibits the weakest performance, potentially attributed to ConvLSTM's failure to fully leverage the spatial-temporal features of satellite cloud images. Compared with *S*4, the proposed method improves the ability to capture temporal features by introducing BDR mechanisms into LSTM and SA-ConvLSTM, resulting in significant performance differences between these methods. In addition, by comparing with *S*3, *S*4, and *S*5, it can be shown that making full use of model complementarity is one of the important ways to improve model performance.

Table 2: Performance of the proposed and competitive methods

Methods	RMSE (kW)	MAE (kW)	MAPE (%)	R^2
S1	19.755	10.732	24.422	0.919
*S*2	20.192	11.872	22.821	0.915
*S*3	18.414	8.748	21.557	0.925
*S*4	19.650	10.039	22.617	0.920
*S*5	18.225	8.429	19.235	0.931
Proposed	**17.782**	**7.753**	**17.926**	**0.935**

3.2.2 Comparisons of Different Horizons

Table 3 displays the evaluation metrics for the proposed method and comparative methods across various time horizons. These methods were assessed over three prediction horizons: 2, 3, and 4 h. In general, the accuracy of predictions tends to diminish as the horizon extends. As shown in Table 3, the R^2 of the proposed method decreases from 0.909 when predicting 2 h ahead to 0.878 when predicting 4 h ahead, primarily due to the insufficiency of historical data to fully capture all changes in future time steps. Despite this decrease, the proposed method consistently outperforms the five competitive methods across all three prediction horizons. Furthermore, it is evident that the performance superiority of the proposed method becomes more pronounced with longer prediction horizons, suggesting that the proposed method exhibits better robustness and stability.

Table 3: Performance of the proposed and competitive methods in different horizons

Horizon	Metrics	*S*1	*S*2	*S*3	*S*4	*S*5	*Proposed*
2-h	RMSE (kW)	25.531	24.126	22.511	23.554	21.924	**20.989**
	MAE (kW)	12.981	13.244	12.015	12.463	12.212	**10.597**
	MAPE (%)	30.404	26.546	25.934	23.581	22.687	**21.967**
	R^2	0.865	0.879	0.895	0.885	0.900	**0.909**
3-h	RMSE (kW)	27.394	26.529	24.530	27.143	25.086	**23.286**
	MAE (kW)	15.271	14.704	13.031	15.298	13.497	**11.864**
	MAPE (%)	33.224	28.431	27.844	30.921	26.993	**24.025**
	R^2	0.845	0.854	0.876	0.848	0.869	**0.888**

(Continued)

Table 3 (continued)

Horizon	Metrics	*S*1	*S*2	*S*3	*S*4	*S*5	*Proposed*
4-h	RMSE (kW)	28.494	26.802	28.369	29.427	26.713	**24.331**
	MAE (kW)	16.289	14.438	16.003	16.627	15.225	**12.641**
	MAPE (%)	35.356	32.966	34.863	36.746	30.558	**26.536**
	R^2	0.832	0.842	0.834	0.821	0.853	**0.878**

3.2.3 Comparisons of Different PV Power Stations

In order to provide additional evidence of the effectiveness of the proposed method, we conducted an evaluation of its performance using the PV power generation data obtained from two supplementary locations within the DKA Solar Centre. Specifically, the analysis included Farm #1 situated at Yulara Service Station, boasting a total power capacity of 226.8 kW, and Farm #2 located at Connellan Airport, with a power capacity of 105.9 kW. Compared to other methods at Farm #1, Table 4 demonstrated that the proposed method exhibited enhancements in RMSE and MAE by 3.3% to 11.5% and 7.6% to 27.2%, respectively. While the proposed method did not surpass all competitive methods at Farm #2, it still demonstrated a competitive performance across the board. Across both PV datasets, the proposed method consistently showed higher predictive accuracy compared to the competitive methods, suggesting that it has superior performance and stronger generalization capabilities.

Table 4: Comparisons between the proposed and competitive methods for different stations

Station	Metrics	*S*1	*S*2	*S*3	*S*4	*S*5	Proposed
#1	RMSE (kW)	9.526	9.351	9.224	9.426	8.713	**8.427**
	MAE (kW)	5.277	5.025	5.240	4.970	4.156	**3.840**
	MAPE (%)	22.664	22.367	21.687	21.134	19.268	**17.867**
	R^2	0.915	0.918	0.920	0.917	0.929	**0.933**
#2	RMSE (kW)	19.923	18.296	17.756	17.243	17.946	**16.953**
	MAE (kW)	11.458	9.386	8.672	**7.804**	9.426	8.502
	MAPE (%)	23.365	21.426	20.129	16.756	20.785	**16.735**
	R^2	0.908	0.922	0.927	0.931	0.925	**0.934**

3.2.4 Comparisons with Other Recent Methods

To substantiate the effectiveness of our proposed method, we compared it with three state-of-the-art deep learning methods. Model 1 (M1) integratess 3DCNN and one-dimensional CNNs to separately capture the spatial-temporal dynamics of cloud formations and the historical PV power generation data [31]. Model (M2) operates by simultaneously analyzing several satellite cloud images, employing normalization, convolutional operations, and attention mechanisms to process the data [32]. Model (M3) utilizes the high-resolution net (HRNet) to preserve high-frequency details in satellite images for more accurate prediction [33].

The overall comparison results for PV power prediction are presented in Table 5. According to the prediction results, the prediction performance of the proposed method is better than the three recent methods. Although recent methods have effectively improved on the traditional methods, the proposed method seems to be able to make better use of the feature in the original data due to the multi-stage temporal feature learning method. Furthermore, the prediction curves for different methods (Fig. 9) indicate that our proposed method exhibits the lowest levels of error. On the clear-sky days, all the methods achieve high accuracy. When considerable weather change is involved, the proposed method exhibits the best prediction accuracy among all methods, which shows that the improved SA-ConvLSTM can better capture the spatial-temporal characteristics of clouds for prediction. By comparing with recent methods, the proposed method is again proven to have competitive performance in the short-term prediction of photovoltaic power.

Table 5: Comparisons between the proposed and other recent methods

Methods	RMSE (kW)	MAE (kW)	MAPE (%)	R^2
M1	21.551	12.174	25.802	0.905
*M*2	18.946	9.832	22.784	0.925
*M*3	22.327	13.302	26.368	0.897
Proposed	**17.782**	**7.753**	**17.926**	**0.935**

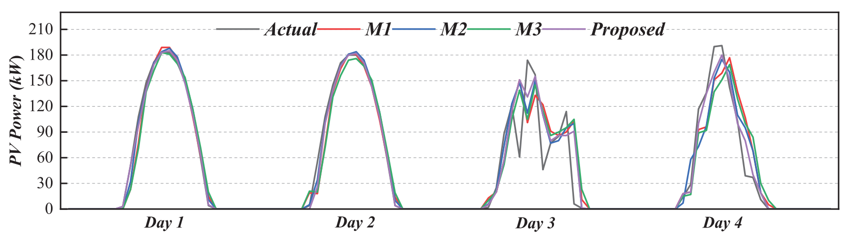

Figure 9: Prediction curves of the proposed and other recent methods

3.3 Discussions of the Proposed Method

3.3.1 Sensitivity Analysis

Selecting the optimal hyperparameters is essential for boosting both the precision and the ability of the proposed method to apply to various situations. We considered several key hyperparameters for each component, including the number of LSTM layers, the number of SA-ConvLSTM layers, and the embedding dimension and number of heads in the Transformer. These optimal hyperparameters were determined through sensitivity analysis on these vital hyperparameters, as illustrated in Fig. 10. The proposed method's performance varied within a small range under different hyperparameter settings, except for the embedding dimension of the Transformer, which demonstrated its strong robustness.

3.3.2 Convergence Analysis

The importance of conducting a convergence analysis is paramount for guaranteeing the stability and dependability of the model training process, as well as for bolstering the model's ability to generalize effectively in practical applications. As depicted in Fig. 11, the training and validation loss of the proposed method demonstrates that it achieves stability by the fifth epoch. Based on the early-stopping strategy (patience = 5), the proposed model finally converged at the 13th epoch, with training and validation losses of 0.0087 and 0.0076, respectively. This rapid convergence confirms the proposed method's efficient and robust learning capability.

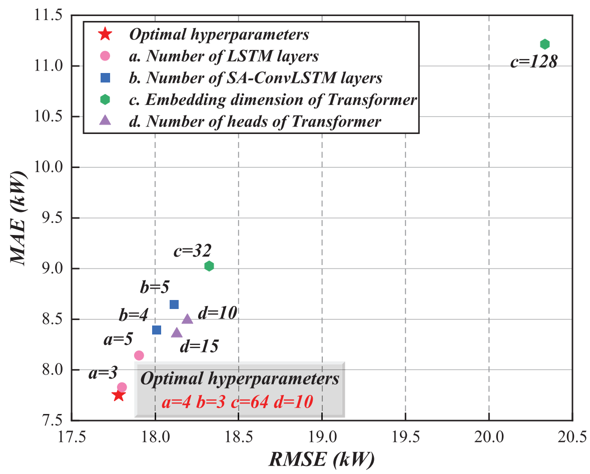

Figure 10: Visualization of the impact of hyperparameters

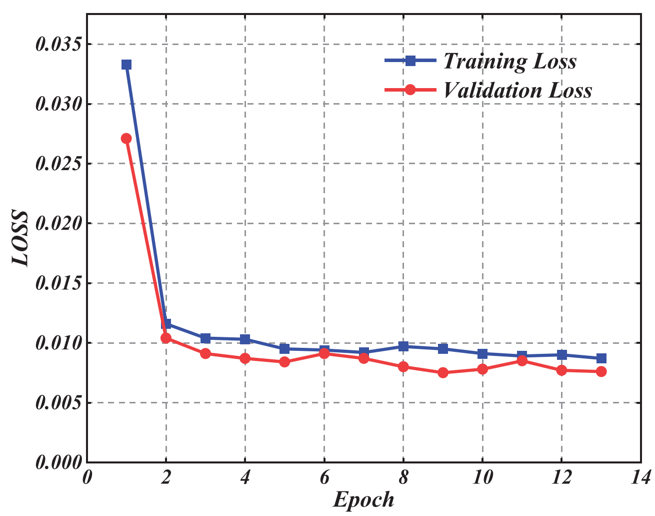

Figure 11: Convergence analysis of the proposed method

3.3.3 Running Time Analysis

The running time of all methods for training and testing is shown in Table 6. Due to the combination of LSTM and SA-ConvLSTM, the proposed method needs more time at each training epoch. Compared with other comparison methods, better learning capability makes the proposed method converge faster, so the total training time is shorter. Given that the prediction duration (7.76 s) remains significantly less than the interval between predictions (60 min), the proposed method proves to be viable for real-time implementation.

Table 6: Comparison of running times for various prediction methods

Method	Training time/epoch (s)	Total training time (s)	Testing time (s)
$S1$	7.65	159.09	6.36
$S2$	9.22	317.99	4.20
$S3$	11.66	244.08	6.43
$S4$	8.82	213.43	6.47
$S5$	12.06	228.06	8.31
Proposed	13.53	180.11	7.76

4 Conclusion

Improving the accuracy of PV power prediction holds significant importance. While existing research has made strides in short-term prediction accuracy, challenges persist due to the underutilization of temporal features and lack of consideration for the relationship between satellite cloud image and PV power generation. These limitations hinder the improvement of PV power prediction performance. Therefore, this study introduces a novel LSTM-Transformer hybrid framework for short-term PV power prediction. The methodology incorporates a bidirectional dynamic residual mechanism in LSTM and SA-ConvLSTM to enhance temporal feature capability and model stability. Furthermore, a novel hybrid attention mechanism is proposed to capture multi-modal complementarities and reduce redundancy. Last, the transformer is applied to further capture the short-term time series patterns and long-term dependencies.

Experiments have demonstrated that the proposed method outperforms competitive methods in short-term prediction. Moreover, its effectiveness in making predictions across multiple steps and locations illustrates its broad generalization and robustness. Nevertheless, there are still obstacles that need to be overcome. For instance, in predicting ultra-short-term outcomes, satellite cloud images may not offer adequate cloud data. One potential solution is to combine both satellite and ground-based cloud images as inputs to enhance predictive accuracy over different time frames. Additionally, the lack of interpretability in deep learning-based forecasting models hinders the ability to analyze their decision-making processes. A promising avenue for enhancing interpretability is to integrate physical principles into these methods.

Acknowledgement: None.

Funding Statement: This work is supported by the Science and Technology Project of Jiangsu Coastal Power Infrastructure Intelligent Engineering Research Center "Photovoltaic Power Prediction System Driven by Deep Learning and Multi-Source Data Fusion" (F2024-5044).

Author Contributions: The authors confirm their contribution to the paper as follows: study conception and design: Qiang Wang, Hao Cheng; data collection, analyzing and interpreting the results, analyzing and interpreting the results: Wenrui Zhang, Guangxi Li; manuscript writing: Fan Xu; data exemplification: Dianhao Chen, Haixiang Zang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The source of the data is the PV dataset from the Data Centre for Photovoltaics and Imagery in Australia (DKASC).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

[1] K. Liu, C. Liang, X. Dong, and L. Liu, "Analysis and modeling of time output characteristics for distributed photovoltaic and energy storage," *Energy Eng.*, vol. 121, no. 4, pp. 933–949, 2024. doi: 10.32604/ee.2023.043658.

[2] P. Li, Z. Wu, C. Zhang, Y. Xu, Z. Dong and M. Hu, "Multi-timescale affinely adjustable robust reactive power dispatch of distribution networks integrated with high penetration of PV," *J. Mod. Power Syst. Clean Energy*, vol. 11, no. 1, pp. 324–334, 2023. doi: 10.35833/MPCE.2020.000624.

[3] W. Ma, "Short-term prediction of photovoltaic power generation based on LMD permutation entropy and singular spectrum analysis," *Energy Eng.*, vol. 120, no. 7, pp. 1685–1699, 2023. doi: 10.32604/ee.2023.025404.

[4] M. Yu *et al.*, "Short-term photovoltaic power point-interval forecasting based on double-layer decomposition and WOA-BiLSTM-Attention and considering weather classification," *Energy*, vol. 275, Jul. 2023, Art. no. 127348. doi: 10.1016/j.energy.2023.127348.

[5] J. Liu, H. Zang, L. Cheng, T. Ding, Z. Wei and G. Sun, "A transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting," *Appl. Energy*, vol. 342, Jul. 2023, Art. no. 121160. doi: 10.1016/j.apenergy.2023.121160.

[6] M. Paulescu *et al.*, "Intra-hour PV power forecasting based on sky imagery," *Energy*, vol. 279, Sep. 2023, Art. no. 128135. doi: 10.1016/j.energy.2023.128135.

[7] Y. Fu *et al.*, "Sky image prediction model based on convolutional auto-encoder for minutely solar PV power forecasting," *IEEE Trans. Ind. Appl.*, vol. 57, no. 4, pp. 3272–3281, 2021. doi: 10.1109/TIA.2021.3072025.

[8] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, "Solar irradiance resource and forecasting: A comprehensive review," *IET Renew. Power Gener.*, vol. 14, no. 10, pp. 1641–1656, 2020. doi: 10.1049/iet-rpg.2019.1227.

[9] H. Zang, D. Chen, J. Liu, L. Cheng, G. Sun and Z. Wei, "Improving ultra-short-term photovoltaic power forecasting using a novel sky-image-based framework considering spatial-temporal feature interaction," *Energy*, vol. 293, Apr. 2024, Art. no. 130538. doi: 10.1016/j.energy.2024.130538.

[10] L. Cheng, H. Zang, Z. Wei, T. Ding, and G. Sun, "Solar power prediction based on satellite measurements—A graphical learning method for tracking cloud motion," *IEEE Trans. Power Syst.*, vol. 37, no. 3, pp. 2335–2345, 2022. doi: 10.1109/TPWRS.2021.3119338.

[11] J. Duan *et al.*, "A multistep short-term solar radiation forecasting model using fully convolutional neural networks and chaotic aquila optimization combining WRF-solar model results," *Energy*, vol. 271, 2023, Art. no. 126980. doi: 10.1016/j.energy.2023.126980.

[12] Y. Zhang, J. Shen, J. Li, X. Yao, X. Chen and D. Liu, "A new lightweight framework based on knowledge distillation for reducing the complexity of multi-modal solar irradiance prediction model," *J. Cleaner Prod.*, vol. 475, 2024, Art. no. 143663. doi: 10.1016/j.jclepro.2024.143663.

[13] X. Huang, J. Liu, S. Xu, C. Li, Q. Li and Y. Tai, "A 3D ConvLSTM-CNN network based on multi-channel color extraction for ultra-short-term solar irradiance forecasting," *Energy*, vol. 272, 2023, Art. no. 127140. doi: 10.1016/j.energy.2023.127140.

[14] J. Yang *et al.*, "Day-Ahead PV power forecasting model based on fine-grained temporal attention and cloud-coverage spatial attention," *IEEE Trans. Sustain. Energy*, vol. 15, no. 2, pp. 1062–1073, 2024. doi: 10.1109/TSTE.2023.3326887.

[15] M. Wang, C. Rao, X. Xiao, Z. Hu, and M. Goh, "Efficient shrinkage temporal convolutional network model for photovoltaic power prediction," *Energy*, vol. 297, 2024, Art. no. 131295. doi: 10.1016/j.energy.2024.131295.

[16] C. Huang and M. Yang, "Memory long and short term time series network for ultra-short-term photovoltaic power forecasting," *Energy*, vol. 279, 2023. doi: 10.1016/j.energy.2023.127961.

[17] Y. Dai, Y. Wang, M. Leng, X. Yang, and Q. Zhou, "LOWESS smoothing and Random Forest based GRU model: A short-term photovoltaic power generation forecasting method," *Energy*, vol. 256, Oct. 2022, Art. no. 124661. doi: 10.1016/j.energy.2022.124661.

[18] M. Abou Houran, S. M. Salman Bukhari, M. H. Zafar, M. Mansoor, and W. Chen, "COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications," *Appl. Energy*, vol. 349, Nov. 2023, Art. no. 121638. doi: 10.1016/j.apenergy.2023.121638.

[19] T. Limouni, R. Yaagoubi, K. Bouziane, K. Guissi, and E. H. Baali, "Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model," *Renew. Energy*, vol. 205, pp. 1010–1024, Mar. 2023. doi: 10.1016/j.renene.2023.01.118.

[20] X. Wang and W. Ma, "A hybrid deep learning model with an optimal strategy based on improved VMD and transformer for short-term photovoltaic power forecasting," *Energy*, vol. 295, May 2024, Art. no. 131071. doi: 10.1016/j.energy.2024.131071.

[21] K. Cao, T. Zhang, and J. Huang, "Advanced hybrid LSTM-transformer architecture for real-time multi-task prediction in engineering systems," *Sci. Rep.*, vol. 14, no. 1, Feb. 2024, Art. no. 4890. doi: 10.1038/s41598-024-55483-x.

[22] Y. Raghuvamsi and K. Teeparthi, "Distribution system state estimation with Transformer-Bi-LSTM-based imputation model for missing measurements," *Neural Comput. Appl.*, vol. 36, no. 3, pp. 1295–1312, 2024. doi: 10.1007/s00521-023-09097-5.

[23] J. Kim, J. Obregon, H. Park, and J. -Y. Jung, "Multi-step photovoltaic power forecasting using transformer and recurrent neural networks," *Renew. Sustain. Energy Rev.*, vol. 200, Aug. 2024, Art. no. 114479. doi: 10.1016/j.rser.2024.114479.

[24] S. Xu, J. Liu, X. Huang, C. Li, Z. Chen and Y. Tai, "Minutely multi-step irradiance forecasting based on all-sky images using LSTM-InformerStack hybrid model with dual feature enhancement," *Renew. Energy*, vol. 224, Apr. 2024, Art. no. 120135. doi: 10.1016/j.renene.2024.120135.

[25] S. Liu, Z. Kong, T. Huang, Y. Du, and W. Xiang, "An ADMM-LSTM framework for short-term load forecasting," *Neural Netw.*, vol. 173, 2024. doi: 10.1016/j.neunet.2024.106150.

[26] R. Zhang, H. Ma, T. K. Saha, and X. Zhou, "Photovoltaic nowcasting with Bi-Level spatio-temporal analysis incorporating sky images," *IEEE Trans. Sustain. Energy*, vol. 12, no. 3, pp. 1766–1776, 2021. doi: 10.1109/TSTE.2021.3064326.

[27] T. Yao *et al.*, "Intra-Hour photovoltaic generation forecasting based on multi-source data and deep learning methods," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 607–618, 2022. doi: 10.1109/TSTE.2021.3123337.

[28] R. Chen, G. Liu, Y. Cao, G. Xiao, and J. Tang, "CGAformer: Multi-scale feature Transformer with MLP architecture for short-term photovoltaic power forecasting," *Energy*, vol. 312, Dec. 2024, Art. no. 133495. doi: 10.1016/j.energy.2024.133495.

[29] S. Lin, S. Wang, X. Xu, R. Li, and P. Shi, "GAOformer: An adaptive spatiotemporal feature fusion transformer utilizing GAT and optimizable graph matrixes for offshore wind speed prediction," *Energy*, vol. 292, Dec. 2024, Art. no. 130404. doi: 10.1016/j.energy.2024.130404.

[30] S. Wang, J. Shi, W. Yang, and Q. Yin, "High and low frequency wind power prediction based on Transformer and BiGRU-Attention," *Energy*, vol. 288, Feb. 2024, Art. no. 129753. doi: 10.1016/j.energy.2023.129753.

[31] J. Qin, H. Jiang, N. Lu, L. Yao, and C. Zhou, "Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning," *Renew. Sustain. Energy Rev.*, vol. 167, 2022. doi: 10.1016/j.rser.2022.112680.

[32] S. Chen, C. Li, Y. Xie, and M. Li, "Global and direct solar irradiance estimation using deep learning and selected spectral satellite images," *Appl. Energy*, vol. 352, 2023. doi: 10.1016/j.apenergy.2023.121979.

[33] H. Ahn, J. Yu, J. Ko, and J. -M. Yeom, "Enhanced short-term prediction of solar radiation using HRNet model with geostationary satellite data," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, 2024. doi: 10.1109/LGRS.2024.3436042.