Applied Soft Computing Journal 178 (2025) 113345

Contents lists available at ScienceDirect

Applied Soft Computing

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

Short-time photovoltaic power forecasting based on Informer model integrating Attention Mechanism

Weijie Yu [a,1], Yeming Dai [a,*], Tao Ren [b,1], Mingming Leng [c]

[a] School of Business, Qingdao University, Qingdao 200071, China
[b] College of Computer Science and Technology, Qingdao University, Qingdao 200071, China
[c] Faculty of Business, Lingnan University, Hong Kong

HIGHLIGHTS

- An AT-Informer-AT model is proposed to predict PV generation.
- Integrating the AM layer to the encoder and decoder for Informer architecture.
- LOWESS smoothing technology is engaged in preprocess the PV data.
- FE technique is utilized to further simplify important features in PV data.
- The prediction efficacy of proposed method is verified on different datasets.

ARTICLE INFO

Keywords:
Photovoltaic power
Informer
Feature engineering
Attention mechanism
Locally weighted scatterplot smoothing

ABSTRACT

Precise Photovoltaic Power Generation Forecasting (PVGF) is significant for achieving reliable power supply, optimizing energy scheduling, and responding to changing energy market demand for sustainable development. However, Photovoltaic Power (PV) is vulnerable to changes in solar radiation levels and temperature, then result in electricity generation fluctuations. To further enhance the precision of PVGF, we propose a new short-term PVGF method based on Informer model integrating attention mechanism. Firstly, Locally Weighted Scatterplot Smoothing (LOWESS) is introduced to preprocess data, enhancing the stability of the input data. Secondly, Feature Engineering (FE) is used for feature screening. Thirdly, Informer model is improved, termed as Attention-Informer-Attention (AT-Informer-AT) model. Specifically, Attention mechanism (AM) layer is added to the encoder and decoder of Informer model respectively, allowing the model to flexibly adjust the attention to different time series data and effectively capture important patterns in the PV data, thereby enhancing prediction performance and generalization ability. Eventually, the novel prediction approach's efficiency is confirmed through analyzing the cases of two different power stations in DKASC area, Alice Springs, Australia and Xuhui District, Shanghai, China. The Experimental results demonstrate that the proposed method superiors other models, with the best prediction accuracy and generalization ability.

1. Introduction

In recent years, sustainable development has become a basic national policy in many countries, then renewable energy, especially solar energy, has increasingly developed in many countries due to its advantages of easy development, cleanliness, harmlessness, and large energy storage capacity. PV, which directly converts solar energy into electricity, is among the most extensively utilized renewable energy generation technologies because of its low development costs, high flexibility, and long service life [1]. Australian government has launched various programs to encourage the adoption of solar power, like the 'Solar School Project' and 'Solar Home and Community Program'. These efforts have led to more than 2 million households installing rooftop solar panels [2]. Similarly, Chinese government has also issued specific policies like "two-carbon policy" to advance the PV's development for low-carbon transitions [3,4].

* Corresponding author.
E-mail addresses: weijieyu212@163.com (W. Yu), yemingdai@163.com (Y. Dai), rentao@qdu.edu.cn (T. Ren), mmleng@ln.edu.hk (M. Leng).
[1] First authors. These authors have equal contributions to the present work.

https://doi.org/10.1016/j.asoc.2025.113345
Received 15 January 2025; Received in revised form 30 April 2025; Accepted 17 May 2025
Available online 23 May 2025
1568-4946/

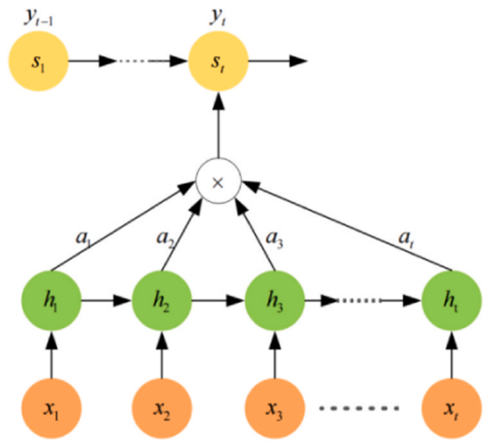


Fig. 1. The framework of Attention Mechanism.

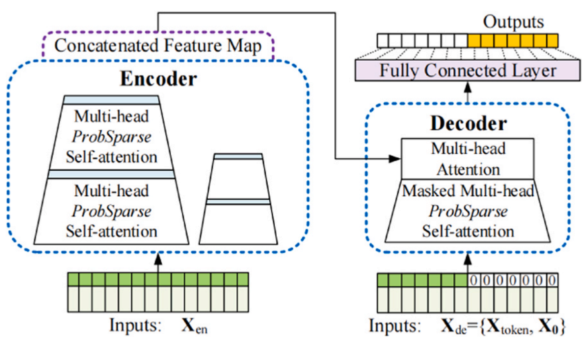


Fig. 2. The structure of Informer model.

As PV is affected by variations in solar radiation intensity and temperature change, accurate PVGF is crucial to improve the PV stations' operational efficacy [5]. Current researches have shown that the accuracy or suitability of PVGF are highly dependent on the forecasting interval and utilized approach. Depending on the forecasting period, PVGF can be grouped into ultra-short-term [6], short-term [7], and medium to long-term [8]. Specifically, ultra-short-term forecasting typically spans from a few seconds to several minutes; short-term forecasting generally covers a horizon from several minutes up to 1–2 days; and medium to long-term forecasting usually refers to time spans ranging from several days or months to multiple years. However, medium to long-term prediction often result in significant errors because of the uncertainty of environmental condition, while ultra-short-term prediction is limited by data collection in a short period of time and may not accurately capture short-term changes. To ensure high PVGF accuracy, it is indispensable to select a more appropriate prediction period [9]. In particular, short-term power prediction not only provides valuable insights for market participants by supporting timely and efficient electricity scheduling, but also plays a critical role in meeting the immediate power demands of PV downstream users. Moreover, it enhances productive interaction with upstream data sources, facilitating real-time decision-making. From a broader perspective, accurate short-term forecasting significantly contributes to the stable and reliable integration of renewable energy into power systems, enabling grid operators to mitigate volatility, optimize dispatch strategies, and support the overall efficiency and resilience of smart grids. However, achieving high accuracy in short-term prediction remains a challenging task due to the inherent intermittency of PV and the complex external influences such as rapidly changing weather conditions. These factors often lead to noisy, incomplete, or inconsistent data, which degrade the quality of model inputs. As a result, accurate short-term forecasting demands not only high-quality data processing, but also the design of robust models to capture intricate temporal dependencies and learn effectively from imperfect information.

At present, PVGF methods are categorized as physical methods [10] as well as data-driven methods [11]. Physical approaches, while intricate and sensitive to environmental conditions, may not be suitable for short-term PVGF [12]. Data-driven methods include statistical and meta-heuristic learning approaches [13]. Statistical approaches contain forecasting derived from time series and Artificial Intelligence (AI) forecasting. Time series prediction [14] can successfully forecast future fluctuations as well as discern patterns, but it often focuses solely on

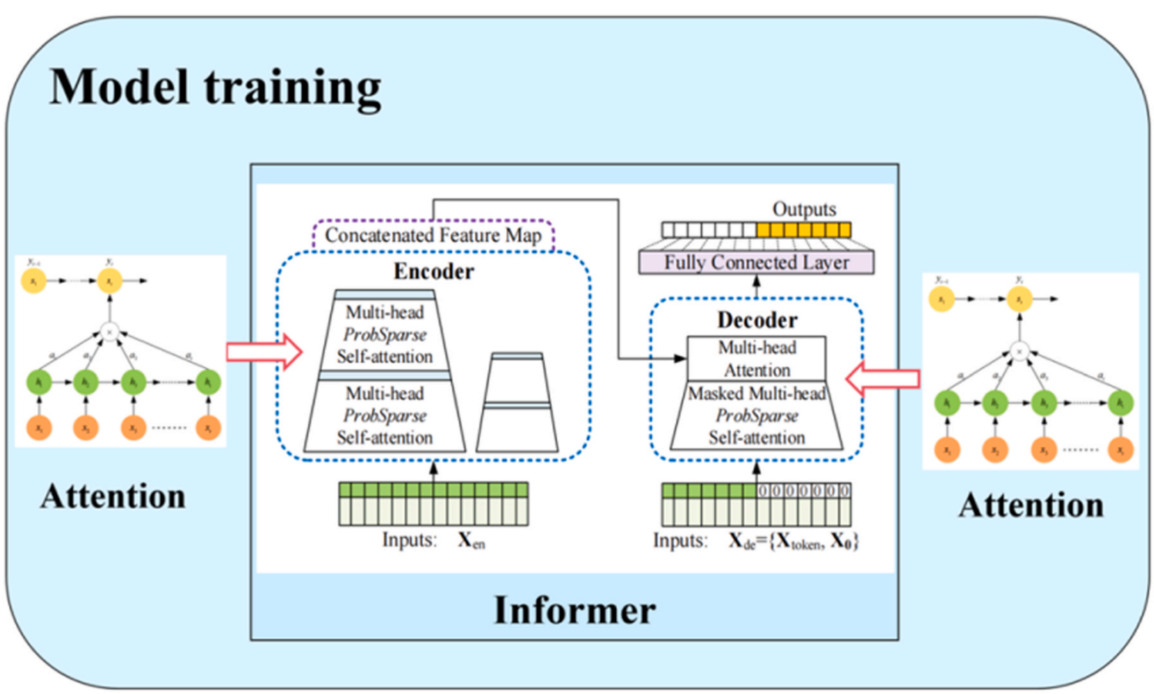


Fig. 3. The framework of AT-Informer-AT model.

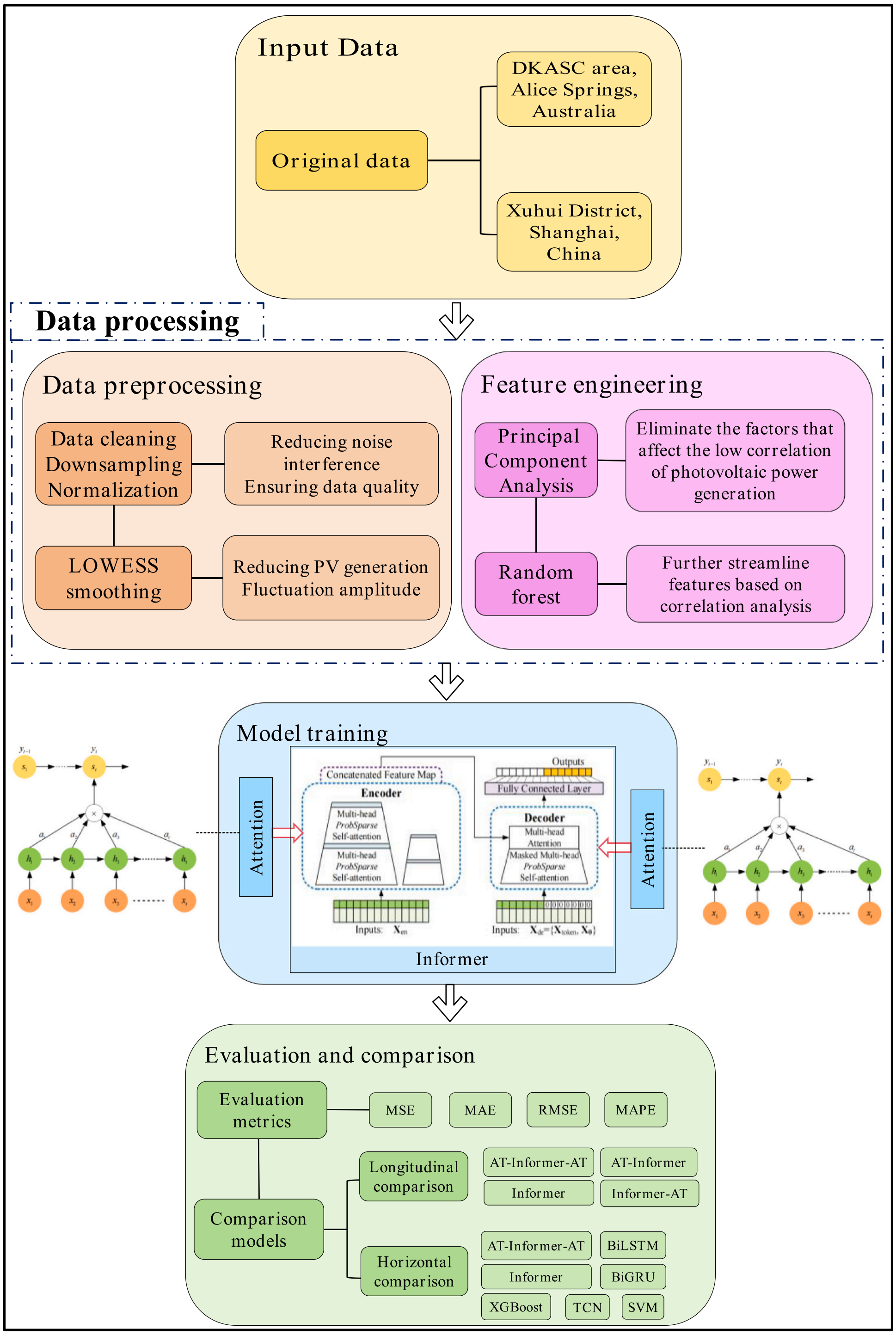
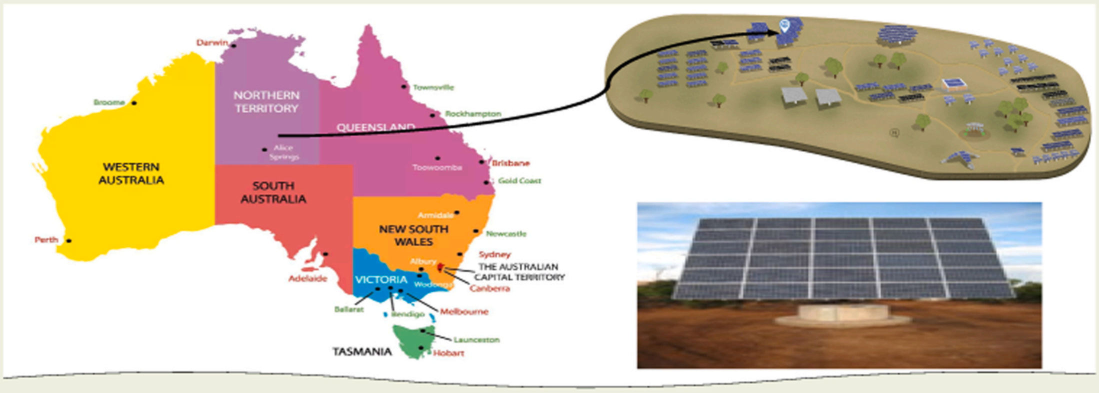


Fig. 4. The overall process of proposed method.

time series data regardless of external elements, leading to high forecasting errors. AI forecasting is dominated by Machine Learning (ML) [15]. ML methods used in load forecasting research mainly include Support Vector Machine (SVM) [16], eXtreme Gradient Boosting (XGBoost) [17], Neural Network (NN) [18]. Among these, NN is widely used in current research because of its capacity for managing vast datasets and enhance model precision [19]. Classical NN include Temporal Convolutional Network (TCN) [20], Bi-directional Long-term Memory (BiLSTM) [21], Bi-directional Gate Recurrent Unit (BiGRU) [22] and Transformer [23]. Among the above, by introducing variable-length convolution kernels and skip connections, TCN performs better than Convolutional Neural Network (CNN) in sequence modeling [24]. BiLSTM demonstrates superior performance over Long-term Memory (LSTM) in sequence data processing, leveraging its bi-directional architecture to effectively assimilate both past and future information [25]. BiGRU is superior to unidirectional Gate Recurrent Unit (GRU) in capturing long-range dependencies in sequences due to its bidirectional gated cyclic unit structure [26]. Transformer, a widely recognized deep learning model, has demonstrated great promise in natural language processing [23]. Nevertheless, its significant time complexity and memory usage constrain its utility in time series prediction. In response to the issues, Zhou et al. [27] put forward the



Array	23.4KW
Panel rating	1.96W
Number of panels	4×30
Panel type	Trina TSM-195DC01A
Array area	4×38.37m2
Type of tracker	DEGER Energie, 5000NT, dual axis
Inverter size/type	4×6KW, SMA, SMC, 6000A
Installation completed	Thu, 8 Jan 2009
Array tilt/azimuth	Variable: Dual axis tracking

Fig. 5. PV station in DKASC.

Designation	Xuhui District Government
Address	Xujiahui in Shanghai, China
PV module capacity	20KW
Grid connection time	2013.3.1
Type of power station	ooftop
Installation type	Fixed bracket
Inverter type	SungrowModbus

Fig. 6. PV station in Xuhui.

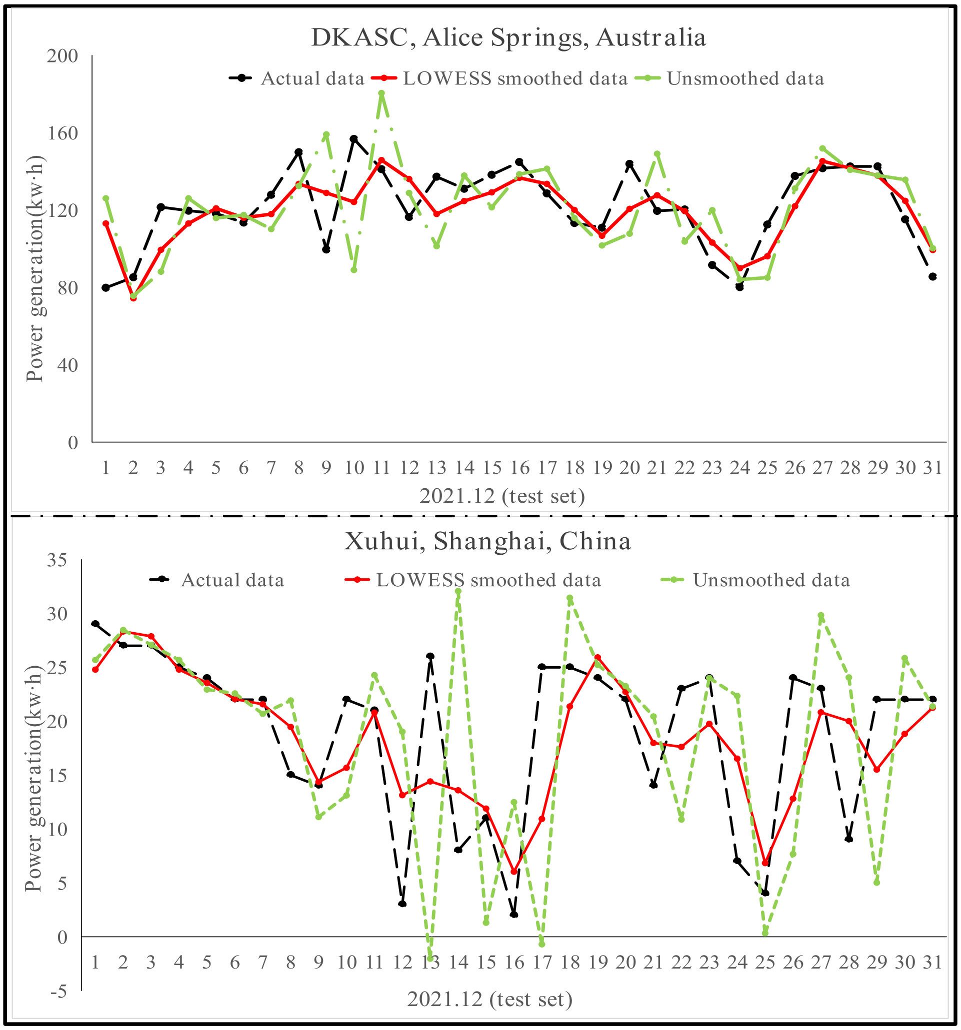


Fig. 7. Display of the smoothing impact.

Informer model as a modification of Transformer, aiming to reduce space complexity and solve the long series forecasting problem. Since its release, Informer has attracted growing attention in the time series domain due to its scalability and effectiveness. Several recent studies have explored its applications in energy-related forecasting tasks, including used Informer model with feature engineering for wind power [28]; utilized Informer model for district heating systems [29]; and respectively incorporated GRU model, LSTM model and AHA-VMD-MPE decomposition reconstruction to improve Informer model for photovoltaic power prediction [30–32]. These works demonstrate the Informer model's strong predictive performance and inspire our application in the context of PVGF. Similarly, the utilization of models grounded in Attention Mechanism (AM) are increasingly gaining popularity in the domain of time series prediction. Du et al. [33] put forward a new temporal AM encoder-decoder model to deal with the challenges of multiple variables time series prediction, showing excellent prediction performance. Ma et al. [34] constructed a hybrid AM-based deep learning method, demonstrating its effectiveness in improving wind power prediction precision. Khan et al. [35] presented a parallel stream network incorporating AM to boost the accuracy of PVGF.

In addition to model prediction, the complete PVGF process also includes data processing, which all impact the forecasting precision [36]. The data processing's essence is to convert original data into an understandable or mineable format to enhance the input data's quality. Nevertheless, during data collection, issues like missing data and outliers frequently arise, which can result in poor-quality mining outcomes and diminish the predictive accuracy of subsequent models. Data processing approaches [37] like smoothing denoising [38] and Feature Engineering (FE) [39] is able to effectively address the aforementioned issues. Due to natural elements like solar radiation and humidity, PV data may exhibit extremely unstable and sporadic. Therefore, it is essential to stabilize the data in research related to PVGF, especially when dealing with substantial data fluctuations. Our earlier research [40] demonstrated that LOWESS smoothing yields optimal smoothing results for PV data, efficiently reducing the seasonal fluctuations of PV and boosting the prediction performance. Moreover, screening important features closely related to PV within input data significantly impacts prediction accuracy [41]. FE enhances model performance and data analysis ability by transforming, combining, and creating new features from the original data. Typically conducted after data preprocessing and before model training, it aims to make data more informative and easier to comprehend by incorporating two main modules: Feature Construction (FC) and Feature Selection (FS) [28]. As a basic tool for classification or regression accuracy, Random Forest (RF) is widely utilized in FS [42]. In [43], the combination of RF and FE is used to predict emissions of nitrogen oxides and carbon dioxide from a trial dual-fuel engine, enhancing the effectiveness and dependability within the modeling system. Therefore, it is vital to process the data to enhance the accuracy of PVGF.

In summary, to further enhance the precision of PVGF, given existing studies, we propose a novel power load prediction method by combining

Table 1
Catagory and Specificity of influential elements.

Category	Specificity
Category 1. Solar Fluxes and Related	(1) All Sky Surface Shortwave Downward Irradiance
	(2) Clear Sky Surface Shortwave Downward Irradiance
	(3) All Sky Insolation Clearness Index
	(4) All Sky Surface Longwave Downward Irradiance (thermal infrared)
	(5) All Sky Surface Photosynthetically Active Radiation (PAR) Total
	(6) Clear Sky Surface Photosynthetically Active Radiation (PAR) Total
	(7) All Sky Surface UVA Irradiance
	(8) All Sky Surface UVB Irradiance
	(9) All Sky Surface UV Index
Category 2. Parameters for Solar Cooking	(1) All Sky Surface Shortwave Downward Irradiance
	(2) Clear Sky Surface Shortwave Downward Irradiance
	(3) Wind Speed at 2 Meters
Category 3. Temperatures	(1) Temperature at 2 Meters
	(2) Dew/ Frost Point at 2 Meters
	(3) Wet bulb Temperature at 2 Meters
	(4) Earth Skin Temperature
	(5) Temperature at 2 Meters Range
	(6) Temperature at 2 Meters Maximum
	(7) Temperature at 2 Meters Minimum
Category 4. Humidity/ Precipitation	(1) Specific Humidity at 2 Meters
	(2) Relative Humidity at 2 Meters
	(3) Precipitation
Category 5. Wind/ Pressure	(1) Surface pressure
	(2) Wind Speed at 10 Meters
	(3) Wind Speed at 10 Meters Maximum
	(4) Wind Speed at 10 Meters Minimum
	(5) Wind Speed at 10 Meters Range
	(6) Wind Direction at 10 Meters
	(7) Wind Speed at 50 Meters
	(8) Wind Speed at 50 Meters Maximum
	(9) Wind Speed at 50 Meters Minimum
	(10) Wind Speed at 50 Meters Range
	(11) Wind Direction at 50 Meters

Table 2
PCA analysis results.

PCA	DKASC		Xuhui	
$\lvert Value \rvert \geq 0.4$	A(1): −0.535	C(7): 0.671	A(2): 0.598	C(2): 0.773
	A(3): −0.651	D(1): 0.791	A(3): 0.628	C(3): 0.829
	A(4): 0.776	D(2): 0.793	A(5): 0.627	C(4): 0.830
	A(5): −0.49	D(3): 0.443	A(6): 0.628	C(5): 0.429
	A(7): −0.447	E(1): −0.477	A(7): 0.628	C(6): 0.855
	B(3): 0.460	E(3): 0.680	A(8): 0.628	C(7): 0.767
	C(1): 0.611	E(4): 0.579	A(9): 0.610	D(1): 0.785
	C(2): 0.672	E(5): 0.803	C(1): 0.862	E(1): −0.430
	C(3): 0.561	E(8): −0.443		
	C(4): 0.613	E(9): 0.581		
	C(5): 0.460	E(10): 0.625		

data processing and optimization of the Informer model. We firstly opt for LOWESS smoothing to preprocess the PV data, and employ FE to further simplify the PV data to comprehensively boost the input data's quality. Then, we add the AM layer to the encoder and decoder of Informer model to improve the model's prediction performance and generalization ability. Finally, through the longitudinal and horizontal comparison analysis based on two datasets from different regions across four seasons, we conclude that the proposed method significantly enhances the performance of PVGF, showing excellent forecasting accuracy and stability. The primary contributions of the paper can be condensed as listed:

Table 3
RF further feature selection results.

District	PCA + RF (RF≥0.1)	No PCA + RF (RF≥0.05)
DKASC	A(3): 0.33	A(1): 0.05
	A(4): 0.45	A(5): 0.05
	C(2): 0.35	A(8): 0.08
	C(7): 0.35	
	D(1): 0.47	
	D(2): 0.47	
	E(3): 0.36	
	E(5): 0.12	
Xuhui	A(2): 0.30	A(1): 0.56
	A(6): 0.27	A(2): 0.10
	A(8): 0.10	A(6): 0.03
	C(2): 0.45	C(5): 0.06
	C(6): 0.13	
	D(1): 0.35	

Table 4
Parameter settings of RF.

Parameter	Parameter meaning	The parameter value
n_estimators	the number of decision trees	200
max_depth	the maximum depth of decision trees	3
random_state	the setting of random numbers	42
min_samples_leaf	the minimum sample number of leaf nodes.	2

Table 5
Parameter settings of Informer model.

Parameter	Parameter meaning	The parameter value
enc_in	Encoder input size	32
dec_in	Decoder input size	32
dec_out	Decoder output size	1
enc_layers	Layers of encoder.	2
dec_layers	Layers of decoder.	1
patience	Patience	3
n_heads	Number of heads	8
batch_size	Batch size	32
learning_rate	Learning rate	0.0001
train_eopchs	Epochs	6
dropout	Dropout rate	0.01
optimizer	Optimizer	Adam
loss	Loss functiom	MSE

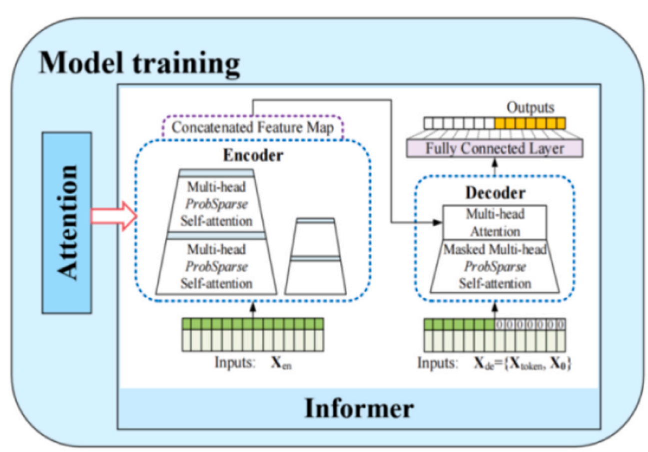


Fig. 8. The AT-Informer model's framework.

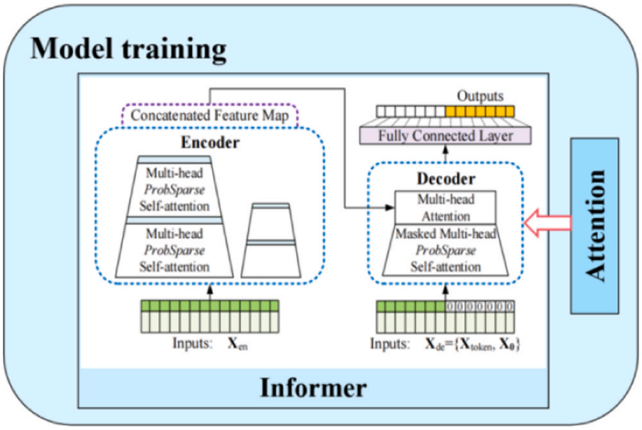


Fig. 9. The Informer-AT model's framework.

(1) We propose a novel AT-Informer-AT model that enhances the Informer architecture by integrating the AM layer to the encoder and decoder respectively.
(2) The FE technique is utilized to further simplify important features in PV data and improve the input data's quality.
(3) The effect of integrating the AM layer to Informer model is verified by ablation experiments.
(4) The prediction efficacy of proposed method is demonstrated through comparison with other models, using two different PV generation datasets from the DKASC in Alice Springs, Australia and Xuhui District in Shanghai, China. These datasets are divided into four seasons for thorough evaluation.

This paper is structured as follows: Section 2 outlines the fundamental principles of LOWESS smoothing approach, Feature engineering, Random Forest, Attention Mechanism and the original Informer model. Section 3 outlines the framework of Informer model that is optimized by AM layer and the detailed process for our forecasting approach. In Section 4, we show the case study, including an account of the empirical data, the data processing results, the model's parameters setting, the selection of evaluation indicators, and a comprehensive analysis of the experimental results including longitudinal comparison and horizontal comparison. Section 5 presents an overview of our work.

2. Theoretical basis

This part outlines the basic methodologies for data processing as well as forecasting. In data processing, methods used involve LOWESS smoothing, Principal Component Analysis and Random Forest. In forecasting, we employ Informer model as the foundational forecasting model.

2.1. LOWESS smoothing

LOWESS smoothing is a non-parametric regression approach applicable for forecasting as well as smoothing. It entails choosing a point x, conducting a weighted linear regression on data points in a specific proximity of x. The regression weights are dictated through a weighting function W, typically represented as a cubic function as depicted in Eq.

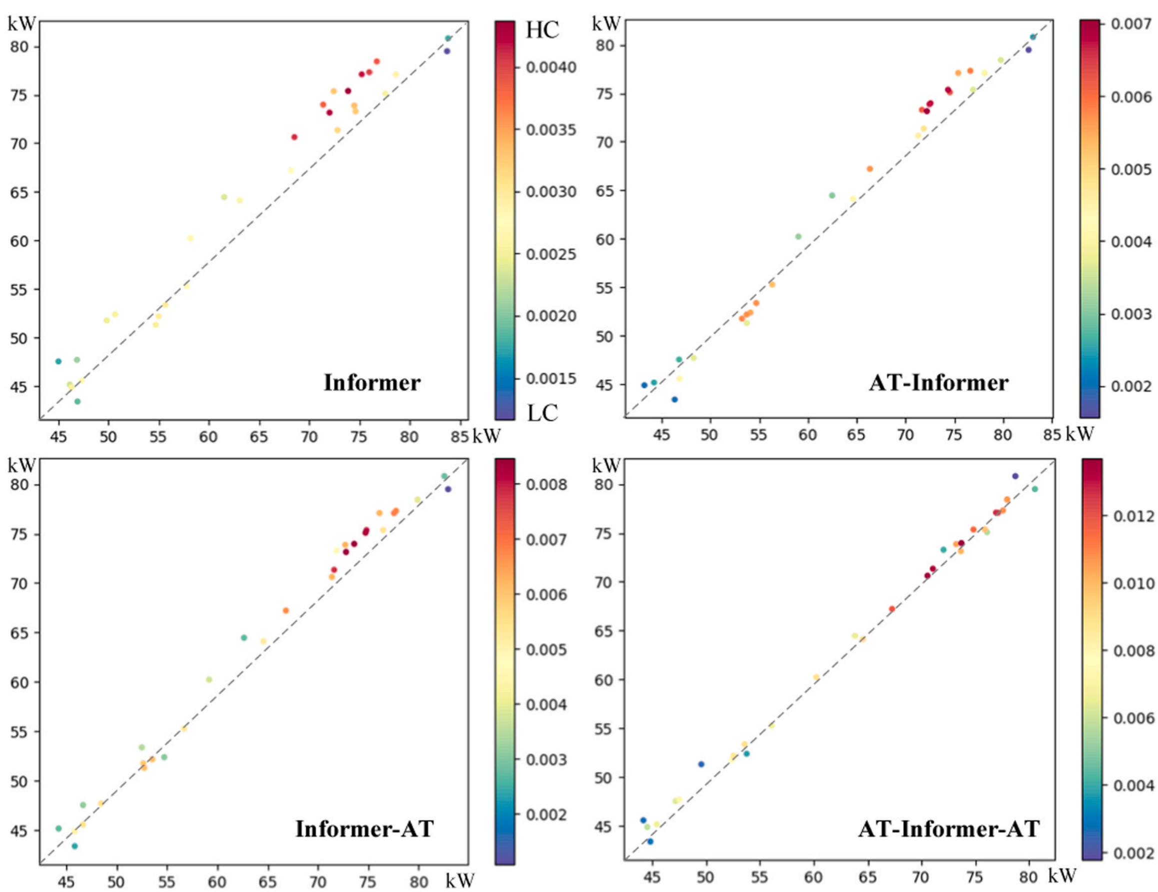


Fig. 10. Density map of actual data compared with prediction results (Spring in DKASC). Annotation: "kW" is the unit for both axes, indicating that the PV value are measured in kilowatts. "HC" refers to High Concentration and "LC" refers to Low Concentration. The redder the color, the higher the density of data points in that region, which reflects a stronger agreement between the predicted and actual values. Figs. 11–13 follow the same design principle.

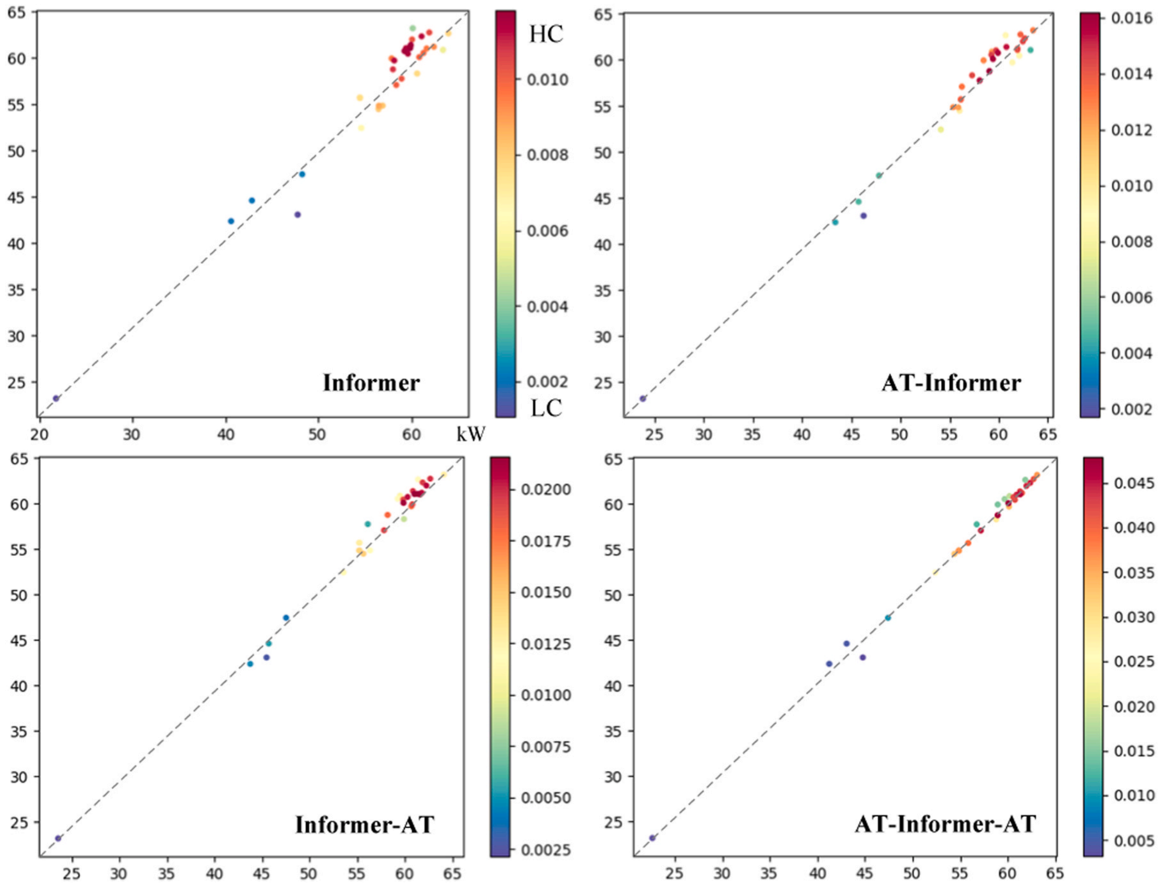


Fig. 11. Density map of actual data compared with prediction results (Summer in DKASC).

(1). The procedure is iterated for all n data points, yielding n weighted regression lines [44].

$$W_{(x)} = \begin{cases} (1-|x|)^3, if|x| < 0 \\ 0, if|x| \geq 0 \end{cases} \tag{1}$$

2.2. Feature Engineering

FE is a crucial step in ML and data mining, serving as the support of numerous analytical approaches that have the potential to greatly enhance model's efficacy [41]. The primary objective of FE is to identify and extract distinct characteristics relevant to the target variables from the initial dataset. These extracted features are subsequently utilized to refine the predictive model, pushing it towards maximizing its feature determination capabilities.

FE is composed of two core modules: FC and FS. In FC, new features are created to enhance the predictive power of the data, including nonlinear transformation, Principal Component Analysis (PCA), and so on [45]. PCA is a renowned technique utilized for both compressing data and extracting features. Its excellence lies in its ability to compute the eigenvectors of the initial inputs' covariance matrix through linear transformation of a vector with high dimensions. Furthermore, FS involves eliminating superfluous or inconsequential characteristics from the set following FC. This process also helps to extract valuable feature subsets, resulting in enhanced model efficacy and forecasting precision [28].

2.3. Random Forest

RF technique, employing random resampling bootstrap and node random classification, constructs numerous decision trees independently. Ultimately, classification results are determined through a voting process [42]. Assessing variable importance can serve as a FS mechanism for high-dimensional data, which is helpful to analyze the importance of factors affecting PV. The equation for evaluating feature significance with RF is provided as Eq. (2). Within this formula, VI represents the significance of data, $error1$ denotes the out-of-bag error from building a RF model using in-bag data, $error2$ reflects the out-of-bag error attained by randomly altering specific data units in out-of-bag data fragments, and N represents the number of RF decision trees. The introduction of random noise leads to an increase in$error2$, implying the significance of the feature and its substantial impact on forecasting outcomes.

$$VI = \sum (error2 - error1) \Big/ N \tag{2}$$

2.4. Attention Mechanism

AM is a model which simulates the attention of human brain. The ideology comes from the idea that human mind directs its attention toward matters at a particular moment to a specific place, so as to reduce or ignore the attention of other parts. By subtly and rationally changing the focus on information, disregarding irrelevant details, and amplifying key data, it enhances the efficient extraction of required information [23].

The formulas for calculating the weight coefficient in AM are

$$e_i = v_a^T \tanh(W_i h_i + b_i) \tag{3}$$

$$a_i = \frac{\exp(e_i)}{\sum_{i=1}^{t} \exp(e_i)} \tag{4}$$

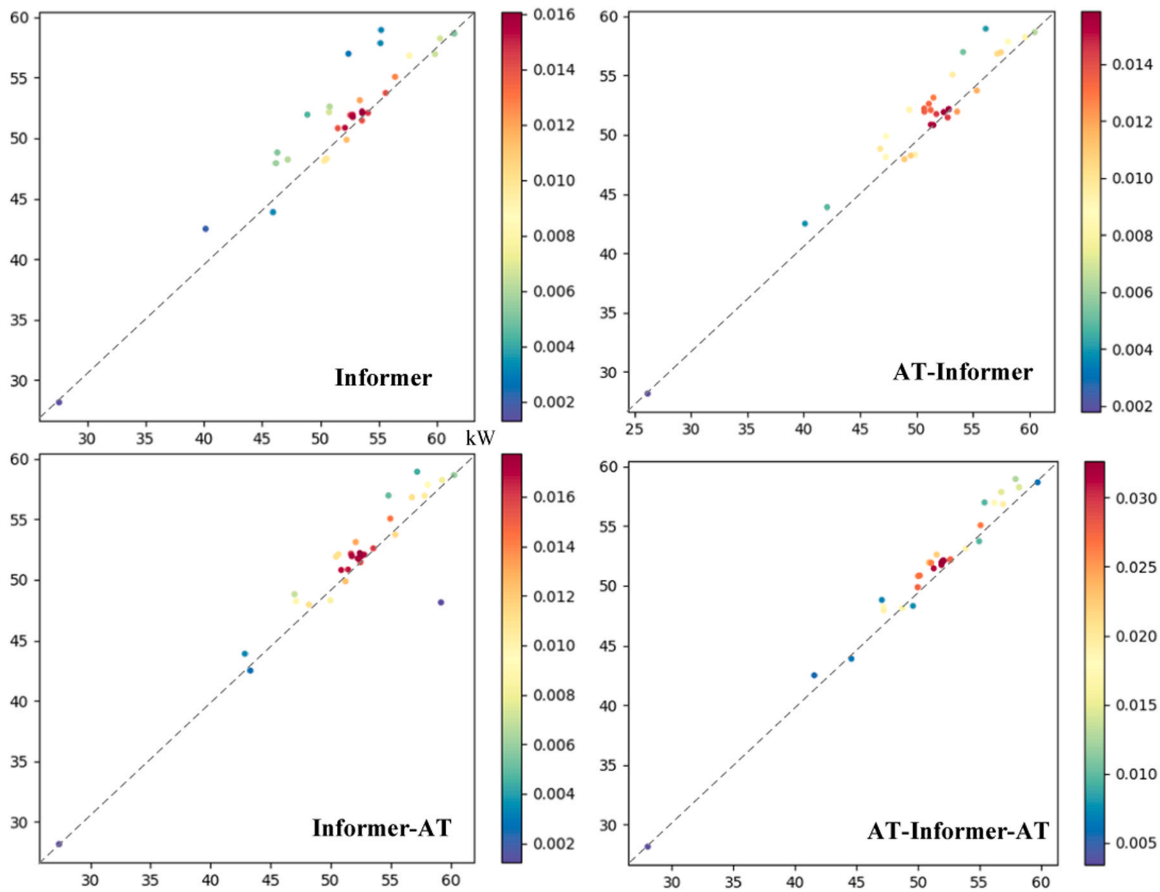


Fig. 12. Density map of actual data compared with prediction results (Autumn in DKASC).

$$s_i = \sum_{i=1}^{t} a_i h_i \tag{5}$$

where e_i refers to the information contained in the hidden layer state h_i, with v_a^T as well as W_i representing the weights, b_i denoting the constant term, and s_i signifying the attention layer output at time t. The AM framework is presented in Fig. 1.

2.5. Informer model

Informer is a NN model for time series prediction. It performs well in processing multi-scale time series data and has efficient computational performance. Its structure is demonstrated in Fig. 2. The Informer model adopts a called Encoder-Decoder framework, and uses self-attention mechanism and convolution mechanism to capture correlations as well as local patterns within time series data, thus enabling the model to learn and predict time series data more effectively [27]. Among them, the self-attention mechanism calculates the inner product of Q as well as Kperforms *soft* max normalization to obtain attention weights, and then applies these weights to the value matrix to generate a weighted sum. It is weighted according to the importance of different parts of the sequence while retaining the structural information of the input sequence. The specific equation is as listed:

$$Attention\left(Q, K, V\right) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where Q, K, and V denote the matrix of Query, Key, and Value, and d_k refers to the key's dimension.

3. Improved Informer prediction method integrating AM

3.1. Informer prediction model integrating AM

The improved Informer model is to add an AM layer in the encoder as well as decoder respectively, denoted as Attention-Informer-Attention (AT-Informer-AT) model, shown in Fig. 3. Firstly, AM layer is added to the encoder to dynamically adjust the attention to different time steps in the input sequence, which make the encoder can more accurately understand the content and input sequence's structure, thereby improving its capacity to catch the dependence and important information across various time steps within the input sequence. Secondly, AM layer is added to the decoder to make better use of the information output by the encoder to dynamically adjust the attention of the decoder at each time step, generating the target sequence more accurately.

By integrating the AM layers into the encoder as well as decoder respectively, the improved Informer model effectively enhances its predictive performance as well as generalizing capacity. Specifically, the AM layer in the encoder dynamically adjusts attention across different time steps in the input sequence, enabling the model to better capture structural patterns and long-range temporal dependencies. Moreover, in the decoder, the AM layer further refines the model's focus on the encoder's output, allowing for dynamic attention modulation at each decoding step. This leads to more accurate and context-aware generation of the target sequence.

3.2. Specific steps of forecasting method

Fig. 4 illustrates the complete flow diagram of the PV load prediction method in accordance with the prediction model introduced in Section 3.1. The prediction procedure is segmented into the subsequent four key

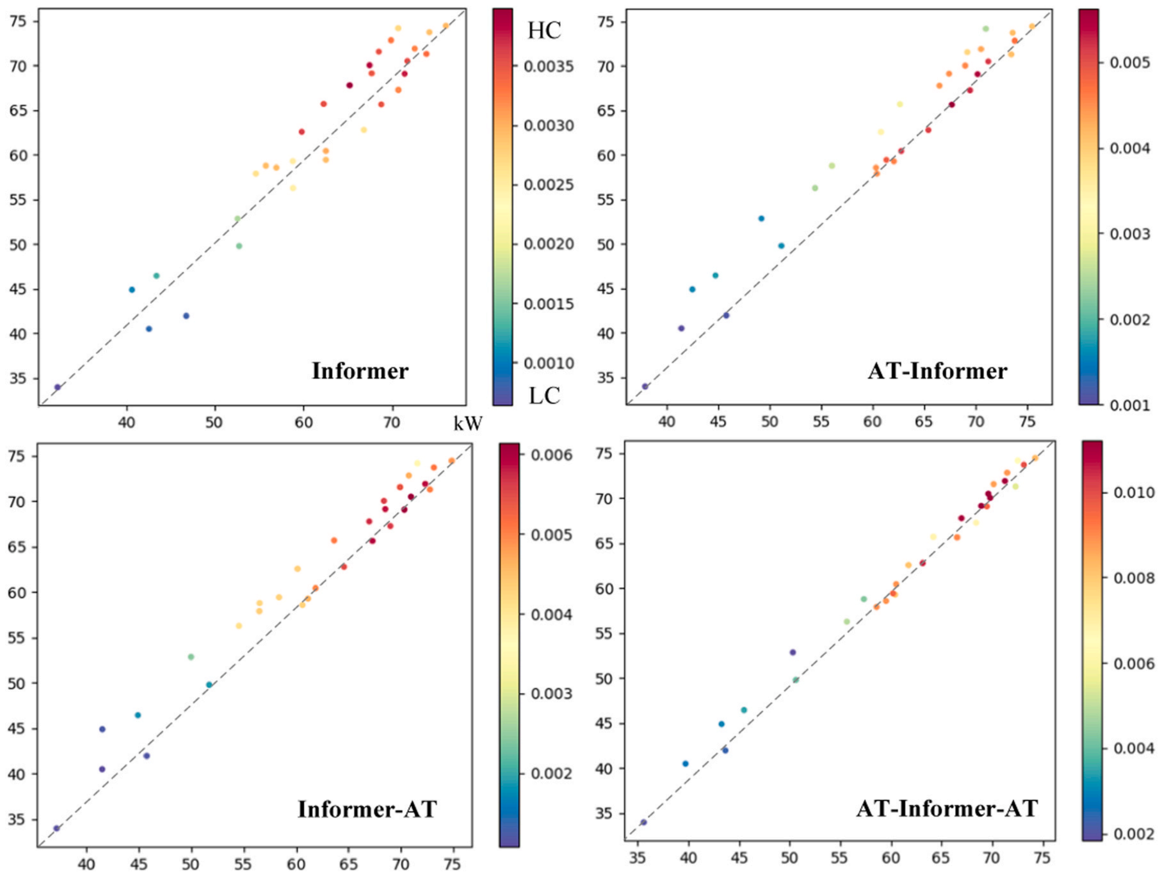


Fig. 13. Density map of actual data compared with prediction results (Winter in DKASC).

stages:

Step. 1 Data preprocessing. Firstly, the original PV data is cleaned, down-sampled, and max-min normalized to eliminate missing values, outliers, and erroneous data, thereby reducing noise interference and ensuring data quality. Secondly, the LOWESS smoothing with the best effect is used to process the actual PV supply data, decrease the daily PV fluctuation range as well as boost the data consistency.

Step. 2 Feature engineering. Firstly, in FC, the PCA with the greatest applicableness is used to compress the smoothed PV supply data, eliminating uncorrelated or very weakly correlated features. Secondly, in FS, the residual features after PCA are further screened by RF to simplify the important features and enhance the input data's quality.

Step. 3 Model training. The AT-Informer-AT forecasting model is constructed. According to the Informer model, the addition of the AM layer to both the encoder and decoder enables the model to dynamically adapt its attention across various time steps in the input sequence. This effectively boots the model's capability to catch important patterns in the PV load sequence data, consequently improving the model's capacity to extrapolate as well as forecast.

Step. 4 Evaluation and comparison of forecasting results. The forecasting outcomes are evaluated with MSE, RMSE, MAPE and MAE, and the method's efficacy corroborated through comparison with other models. The trial outcomes indicate that this method yields the least forecasting error, and lead to the best forecasting effect.

4. Case study

4.1. Data preparation

For securing the comprehensiveness and impartiality of data, we gather datasets from two PV stations, i.e., DKASC (133.52 W, 24.17 S) in Alice Springs, Australia [46] and Xuhui District Government (121.43 E, 31.18 N) in Shanghai, China [47]. These two cities are selected for conducive weather conditions as well as plenty of solar radiation, along with analogous and comparable latitudes. Alice Springs is situated within the Southern and Western Hemispheres, while Shanghai is located within the Northern and Eastern Hemispheres. These cities' amalgamations from diverse hemispheres render them more emblematic.

To fully evaluate our method's efficacy, the power data of two power stations from 2015 to 2022 is selected, and divided the data into training, testing and verification sets in a 7: 1: 2 ratio. Figs. 5 and 6 present detailed information regarding the PV data's positions, PV plants' conditions, and the detailed partitioning of training as well as testing datasets. The data frequency is daily.

4.2. Data processing

4.2.1. Data preprocessing

Our previous study [40] has shown that LOWESS can effectively diminish anomalies in data when processing PV data, making the manipulated data smoother, while ensuring the reliability of real data. Therefore, in this experiment, we first use data cleaning, down-sampling and max-min normalization methods to preprocess the original PV data to remove missing values, outliers and error data, reducing noise interference and ensuring data quality. Then the LOWESS smoothing method is used to process the PV data to reduce the daily fluctuation range of photovoltaic and improve the consistency of the data. We collect the data of DKASC in December 2021 and Xuhui in December 2021 to demonstrate the smoothing effect. The results of contrasting the smoothed and unsmoothed values with actual data are presented in Fig. 7. It is evident that the data processed by LOWESS smoothing have a

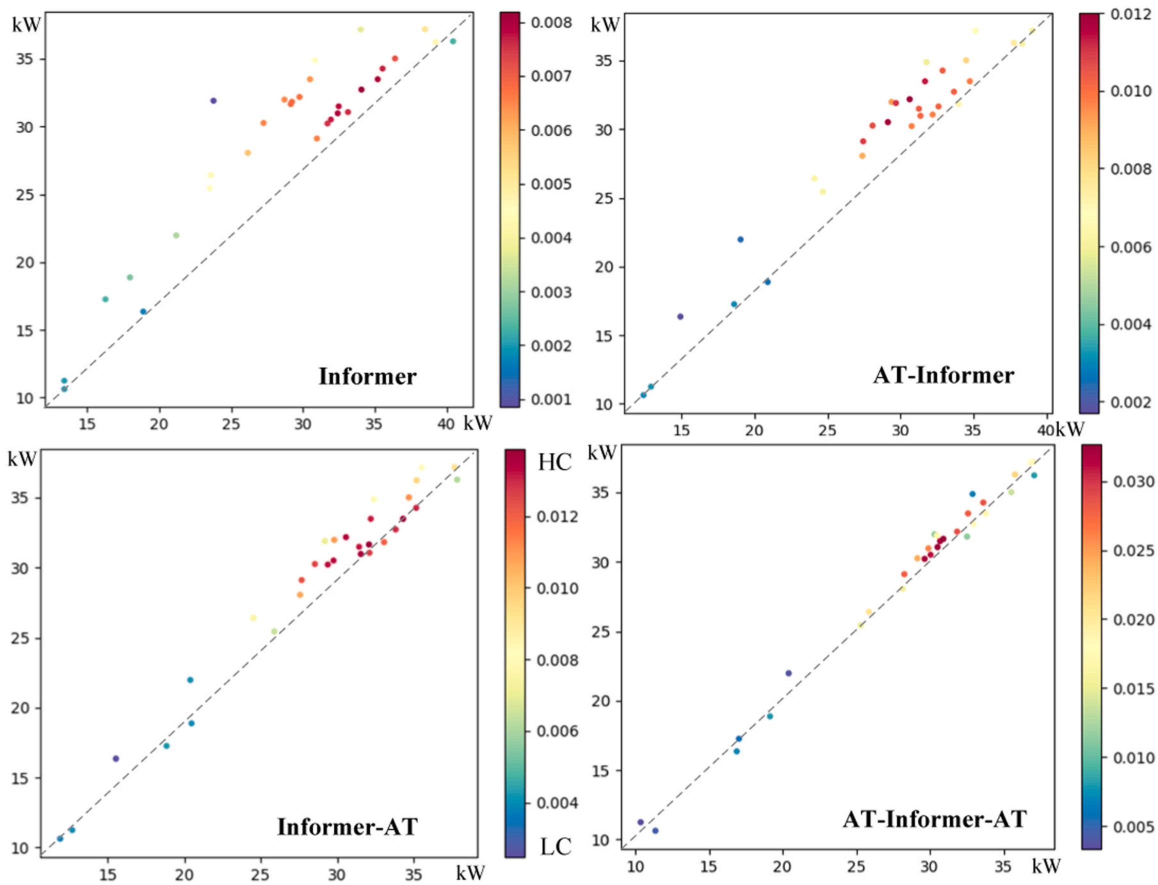


Fig. 14. Density map of actual data compared with prediction results (Spring in Xuhui). Annotation: "kW" is the unit for both axes, indicating that the PV value are measured in kilowatts. "HC" refers to *High Concentration* and "LC" refers to *Low Concentration*. The redder the color, the higher the density of data points in that region, which reflects a stronger agreement between the predicted and actual values. Figs. 15–17 follow the same design principle.

greater degree of fitting within the actual data and a lower fluctuation range. The smoothed prediction results greatly exceed the unsmoothed forecasting outcomes in quality. LOWESS smoothing can significantly enhance the input data's quality by denoising and smoothing the data, boosting the prediction precision, reflecting the necessity for implementing LOWESS smoothing in our study.

4.2.2. *Feature engineering*

(1) Feature construction

To predict PV precisely, it is essential to take into account the diverse influential elements. The researches have pinpointed five primary divisions of aspects [48]: Category 1: Solar Fluxes and Related; Category 2: Parameters for Solar Cooking; Category 3: Temperatures; Category 4: Humidity/ Precipitation; while Category 5: Wind/ Pressure. Table 1 presents a detailed list of the 33 refined influential elements.

Then, we employ PCA to remove uncorrelated and highly weakly correlated characteristics. Table 2 shows the results of various influencing factors after PCA, which absolute value is greater than 0.4.

(2) Feature selection.

After feature construction, RF is used for further refine the remaining features to boost the input data's quality [43]. Table 3 displays the ultimate FS (significance ranking) results after RF filtering.

4.3. *Parameter settings*

All the prediction models in this paper are written on Python and use torch and keras toolkit to build the deep learning framework.

(1) RF parameter settings

The precision of RF model is influenced by several parameters. In accordance with reference [42], we identified the pertinent parameters' values via numerous trials, as presented in Table 4.

(2) Informer model parameter settings

To gain better performance, the parameter setting of Informer model is crucial. Parameter setting involves many aspects, including the input and output settings of the encoder and decoder, layer size, batch size and so on. On the basis of reference [31] and a large number of experiments, we identified the pertinent parameters' values, as presented in Table 5.

4.4. *Evaluation criteria*

In this study, Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) are chosen as the criteria for assessing the forecasting models. y_i and y_i' in Eq. (7)~(10) represent the actual and forecasted values correspondingly.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i'\right| \tag{7}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2 \tag{8}$$

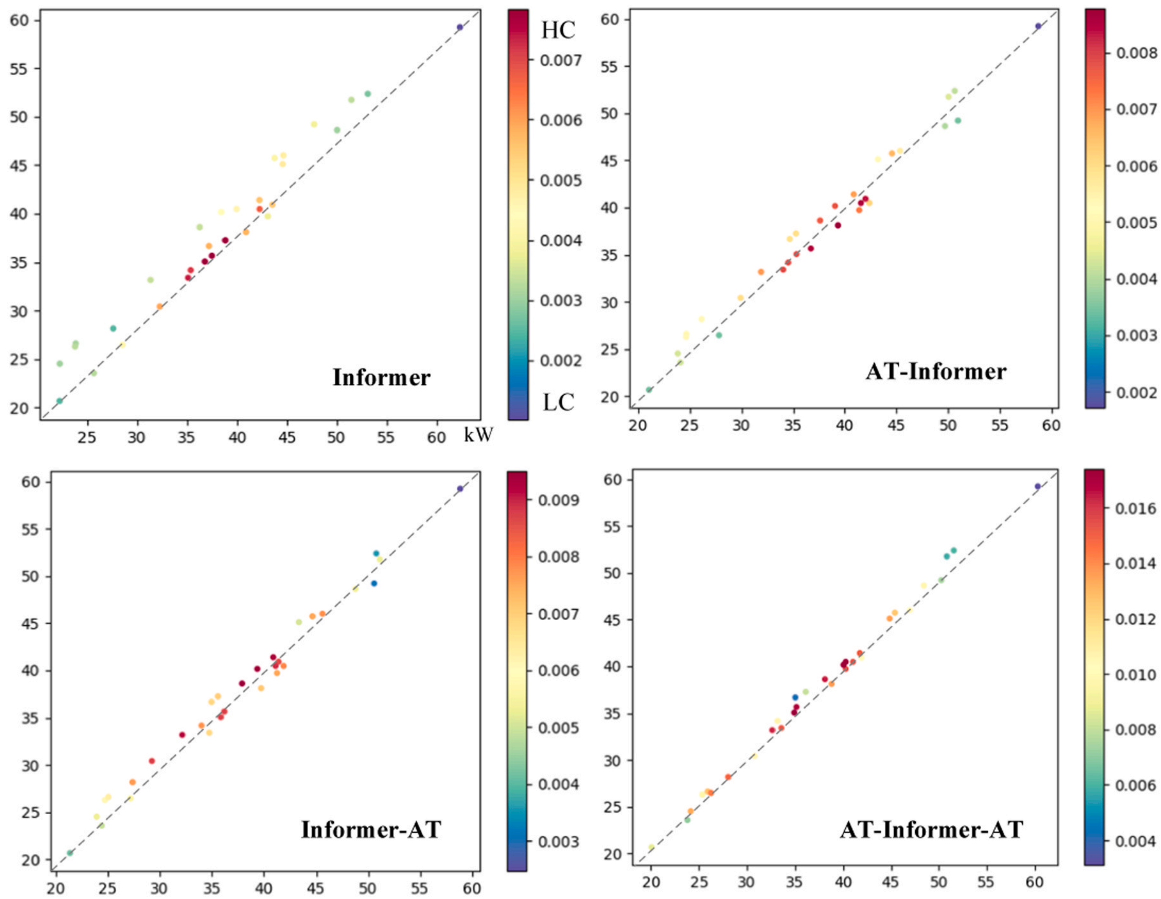


Fig. 15. Density map of actual data compared with prediction results (Summer in Xuhui).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2} \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i'\right|}{y_i} \times 100\% \tag{10}$$

The combination of multiple criteria can comprehensively assess the model's performance. MAE computes the mean of the absolute variance among the true data as well as the forecasted outcomes, which is insensitive to the square error and can better reflect the overall error. MSE can quantify the disparity between the forecasted results and true data, making the performance evaluation of model more intuitive and reliable. RMSE assesses the disparity between forecasted results and actual data, exhibiting heightened sensitivity to outliers and facilitating easier interpretation due to its alignment with the original data units. MAPE intuitively represents the average error of the model relative to actual data and demonstrates greater generality across datasets of varying scales. In conclusion, the utilization of these diverse metrics offers a holistic understanding of a prediction model's performance, taking into account various aspects such as sensitivity to outliers, interpretability, and adaptability to different dataset scales.

4.5. Experiment results and comparative analysis

To thoroughly evaluate the proposed model's efficacy, we conduct both longitudinal comparison and horizontal comparison. For testing across various seasons, we utilize data from DKASC and Xuhui power stations during March, June, September, and December of 2022. Specifically, we designate these months to represent distinct seasons: spring (March of 2022), summer (June of 2022), autumn (September of 2022), and winter (December of 2022). This approach ensures a balanced assessment across different climatic conditions and temporal variations, as each selected month corresponds to the peak characteristics of its respective season, allowing for a comprehensive evaluation of seasonal impacts on photovoltaic power generation. Although DKASC and Xuhui are located in different hemispheres, this seasonal division remains reasonable because it captures the typical weather patterns and solar radiation variations that influence photovoltaic performance in both locations. Therefore, the chosen months effectively represent the seasonal dynamics relevant to each site, enabling meaningful comparisons.

4.5.1. Longitudinal comparison

To investigate the impact of adding AM to the Informer model, particularly the effects of introducing AM separately into the encoder and decoder, we conduct an ablation experiment based on the classic Informer model. Four different model configurations are constructed by adding AM into the model's encoder and decoder components. The specific designs are as follows:

1. Baseline Model: The classic Informer model without any additional AM, as shown in Fig. 2.
2. Attention-Informer (AT-Informer) Model: AM is introduced in the encoder part to improve feature extraction from input sequences, as demonstrated in Fig. 8.
3. Informer-Attention (Informer-AT) Model: AM is introduced in the decoder part to enhance the utilization of contextual information, as presented in Fig. 9.
4. Attention-Informer-Attention (AT-Informer-AT) Model: AM is simultaneously introduced in the model's encoder and decoder parts

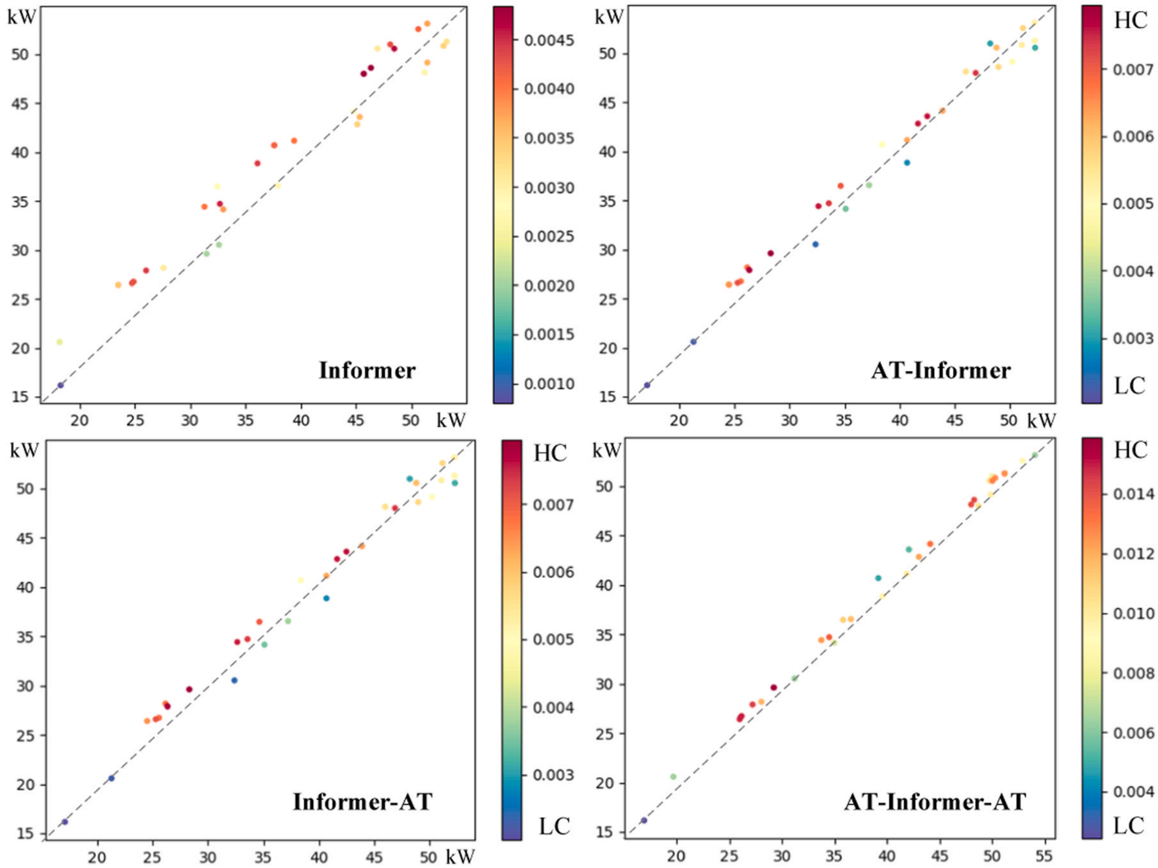


Fig. 16. Density map of actual data compared with prediction results (Autumn in Xuhui).

to catch relevant information from sequences more comprehensively, as displayed in Fig. 3.

To confirm the test results' precision as well as maintain the test conditions' consistency, we employ the same processing methods for the input data across four models. These include the data preprocessing and FE techniques detailed earlier. It ensures that when evaluating the model's performance, the only variable considered is the AM. Through this consistent approach, we can more accurately compare the performance of all models, so as to derive more reliable conclusions.

To compare the correlation between the four models and the actual value more intuitively, we draw the scatter density maps of the corresponding four seasons in the two markets. These are shown in Figs. 10–17. The scatter density map refers to a visual depiction method for visualizing the data distribution. It displays the position and density of scatter points on a plane, representing the data point distribution. Among them, the diagonal line ($y = x$) in the scatter density map represents actual value, while each data point is represented as a scatter point, indicating the model's forecasted value and corresponds to the horizontal and vertical coordinates. In this paper, the two coordinate axes represent the PV value, the unit is kilowatt (kW). Additionally, the color band on the right side of the scatter density map, along with its corresponding values, indicates the connection between the true as well as forecasted values. A dark red color (with large and dense scatters) indicates a High Concentration (HC) between the true as well as forecasted values. Conversely, a dark blue color (with small and sparse scatters) indicates a Low Concentration (LC) between the true as well as forecasted values, the model's precision is reflected by the correlation level. The redder the color, the higher the density of data points in that region, which reflects a stronger agreement between the predicted and actual values. Therefore, by comparing the scatter distribution, we can evaluate the performance and accuracy of a model. The model forecasted outcomes of two cases are as follows:

(1) DKASC area, Alice Springs, Australia

The corresponding scatter density maps of actual data and the four models' forecasted outcomes in spring, summer, autumn and winter are demonstrated in Figs. 10–13, and the error contrast is demonstrated in Table 6. Here is the detailed analysis:

a. Fig. 10 provides a comparison of the four models' effectiveness during the DKASC spring. It is observed that towards the middle-late of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the AT-Informer-AT model displays the densest scatter distribution, indicating a stronger alignment between the true values as well as forecasted values. Conversely, the Informer model demonstrates the sparsest scatter distribution, reflexing a relatively lower prediction accuracy. Based on observation, we can rank the four models' forecasting precision as listed: AT-Informer-AT model > Informer-AT model > AT-Informer model > Informer model.

b. Fig. 11 offers a contrast of the four models' efficacy during the DKASC summer. It is shown that towards the end of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the scatter points of AT-Informer-AT model are basically distributed on the diagonal, indicating a stronger alignment among the forecasted as well as real values. Based on observation, we can rank the four models' forecasting precision as listed: AT-Informer-AT model > AT-Informer model > Informer-AT model > Informer model.

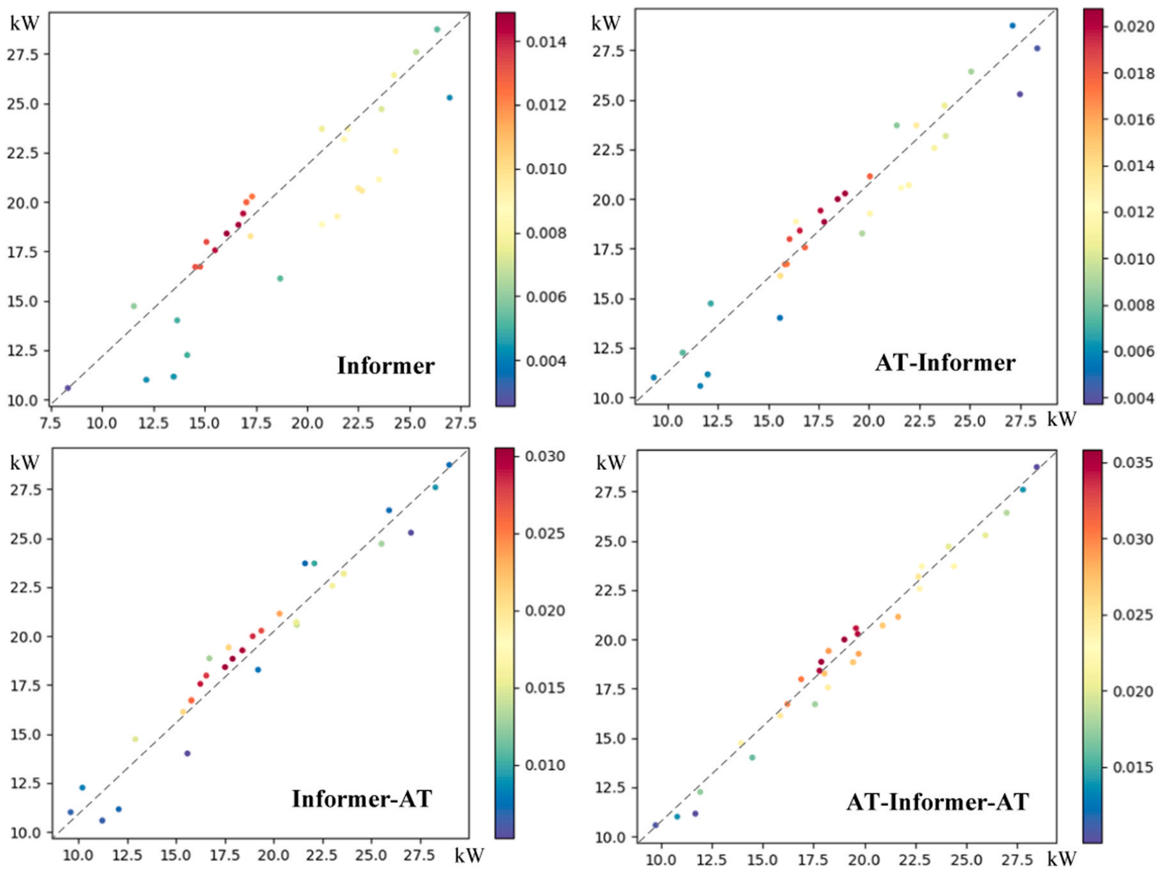


Fig. 17. Density map of actual data compared with prediction results (Winter in Xuhui).

Table 6
The error comparison results of different models (DKASC).

Season	Evaluation criteria	Informer	AT-Informer	Informer-AT	AT-Informer-AT
Spring	MSE	183.3340	181.4427	178.9718	**175.1014**
	MAE	11.0437	11.2104	10.8114	**10.7817**
	RMSE	13.5401	13.4701	13.3780	**13.2326**
	MAPE	10.67 %	10.42 %	10.45 %	**10.40 %**
Summer	MSE	135.0523	133.4163	133.8591	**129.5994**
	MAE	8.5289	**8.2749**	8.3543	8.5475
	RMSE	11.6212	11.5506	11.5697	**11.3842**
	MAPE	18.09 %	17.83 %	17.88 %	**17.70 %**
Autumn	MSE	44.5772	42.7918	43.0167	**39.8330**
	MAE	4.7771	4.5547	4.5883	**4.5141**
	RMSE	6.6766	6.5415	6.5587	**6.3113**
	MAPE	11.45 %	11.00 %	11.03 %	**10.62 %**
Winter	MSE	116.1610	114.5041	117.0791	**111.7468**
	MAE	7.5761	7.3434	7.3770	**7.2971**
	RMSE	10.7778	10.7007	10.8203	**10.5710**
	MAPE	15.54 %	15.40 %	15.70 %	**15.16 %**

Table 7
The error comparison results of different models (Xuhui).

Season	Evaluation criteria	Informer	AT-Informer	Informer-AT	AT-Informer-AT
Spring	MSE	105.6165	103.5143	101.8497	**97.2461**
	MAE	7.9070	7.8833	7.8145	**7.7515**
	RMSE	10.2770	10.1742	10.0921	**9.8613**
	MAPE	12.89 %	12.67 %	12.64 %	**12.33 %**
Summer	MSE	89.1235	87.9204	88.6163	**86.6035**
	MAE	7.8610	7.7335	7.7512	**7.6454**
	RMSE	9.4405	9.3766	9.4136	**9.3061**
	MAPE	9.80 %	9.80 %	9.71 %	**9.54 %**
Autumn	MSE	141.5177	137.5314	130.2626	**129.8167**
	MAE	9.4409	9.4001	9.1771	**9.1766**
	RMSE	11.8961	11.7274	11.4133	**11.3937**
	MAPE	12.68 %	12.28 %	**12.04 %**	12.06 %
Winter	MSE	124.0804	116.4289	115.4726	**114.4745**
	MAE	8.9290	8.7904	8.6322	**8.4072**
	RMSE	11.1391	10.7902	10.7458	**10.6993**
	MAPE	12.20 %	11.56 %	**11.50 %**	11.89 %

c. Fig. 12 demonstrates a comparison of the four models' effectiveness during the DKASC autumn. It is manifested that towards the middle-late of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the AT-Informer-AT model displays the densest scatter distribution, revealing a stronger alignment between the forecasted as well as true values. Conversely, the Informer model demonstrates the sparsest scatter distribution, reflecting a relatively lower prediction accuracy. Based on observation, we can rank the four models' forecasting precision as listed: AT-Informer-AT model > AT-Informer model > Informer-AT model > Informer model.

d. Fig. 13 delivers a contrast of the four models' efficacy during the DKASC winter. It is displayed that compared with the previous three seasons, the scatter distribution of the four models in winter is more dispersed, and the densely distributed scatters are concentrated at the end of the month. Based on observation, we can rank the four models' forecasting precision as listed: AT-Informer-AT model > AT-Informer model > Informer-AT model > Informer model.

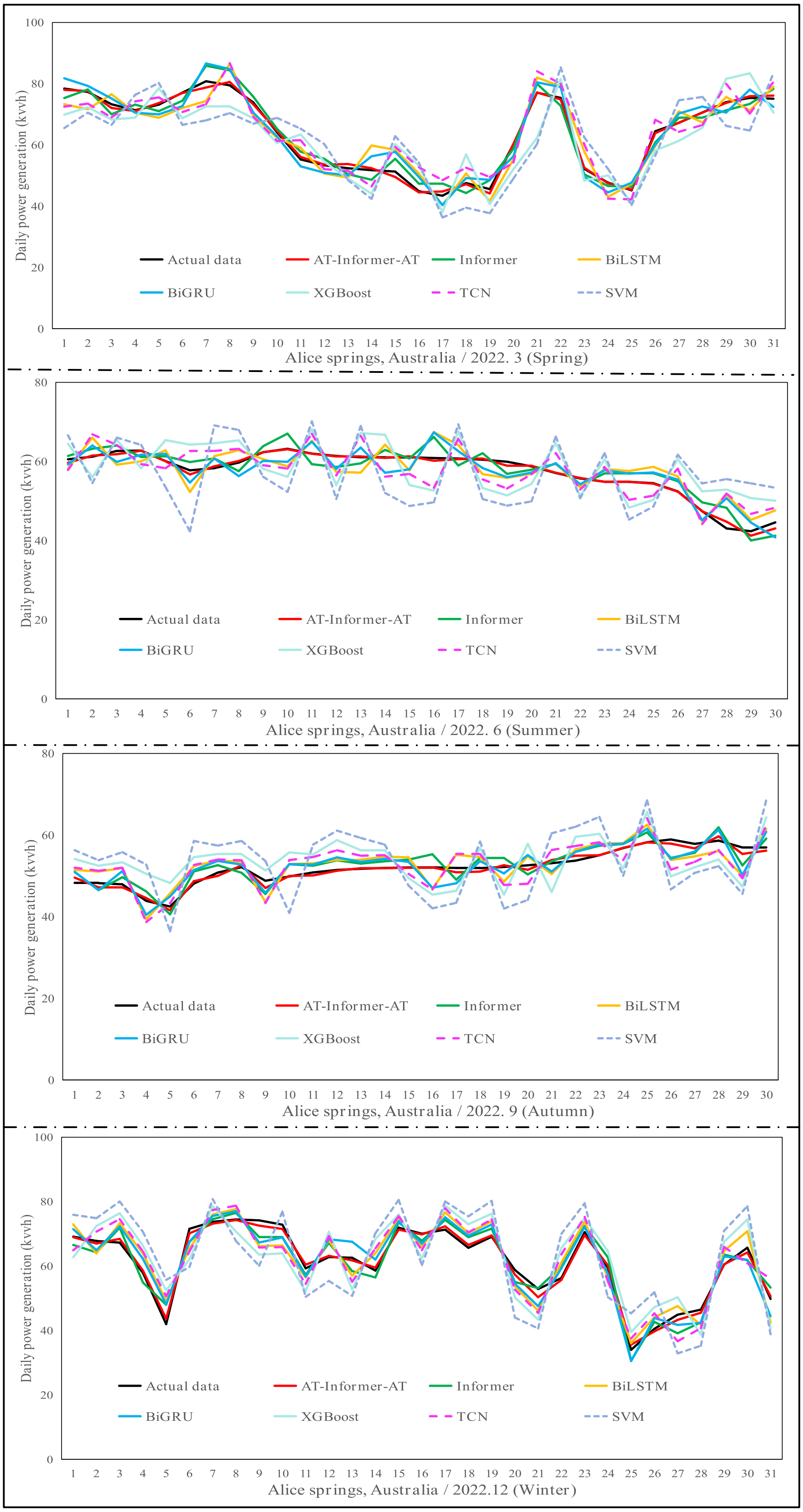


Fig. 18. Comparison between actual data and predicted results (DKASC).

Fig. 19. The error comparison results (DKASC).

e. By comparing the four evaluation criteria of MAE, MSE, RMSE and MAPE in Table 6, it is apparent that in contrast to the other three models, except that the AT-Informer model exhibits the lowest MAE in summer, the other forecasting errors of the AT-Informer-AT model are the smallest across all four seasons, demonstrating that it possesses the best prediction performance and the highest prediction accuracy among the four models.

(2) Xuhui District, Shanghai, China

The corresponding scatter density maps of actual data as well as the four models' forecasting results in spring, summer, autumn and winter are presented in Figs. 14–17, and the error contrast is demonstrated in Table 7. Here is the detailed analysis:

a. Fig. 14 provides a comparison of the four models' effectiveness during the Xuhui spring. It is observed that towards the middle-late of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the AT-Informer-AT model displays the densest scatter distribution, presenting a stronger alignment between the forecasted as well as true values. Conversely, the Informer model demonstrates the sparsest scatter distribution, reflexing a relatively lower prediction accuracy. Based on observation, we can rank the four models' forecasting precision as follows: AT-Informer-AT model > Informer-AT model > AT-Informer model > Informer model.
b. Fig. 15 offers a contrast of the four models' efficacy during the Xuhui summer. It is shown that towards the early middle of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the scatter points of AT-Informer-AT model are basically distributed on the diagonal, indicating a firmer alignment among the forecasted as well as real values. Based on observation, we can rank the four models' forecasting precision as follows: AT-Informer-AT model > AT-Informer model > Informer-AT model > Informer model.
c. Fig. 16 illustrates a comparison of the four models' effectiveness during the Xuhui autumn. It is displayed that compared with the previous two seasons, the scatter distribution of the four models in autumn is more dispersed. Although highly correlated scatters (the dark red color) of the four models are distributed almost throughout the month, the scatter distribution of AT-Informer-AT model is close to the diagonal. Based on observation, we can rank the four models' forecasting precision as follows: AT-Informer-AT model > Informer-AT model > AT-Informer model > Informer model.
d. Fig. 17 delivers a contrast of the four models' efficacy during the Xuhui winter. It is manifested that towards the middle of the month, the predicted values of the four models exhibit a higher correlation with the actual values. Among the models, the AT-Informer-AT model displays the densest scatter distribution, revealing a stronger alignment between the forecasted as well as true values. Conversely, the Informer model demonstrates the sparsest scatter distribution, reflecting a relatively lower prediction accuracy. Based on observation, we can rank the four models' forecasting precision as follows: AT-Informer-AT model > Informer-AT model > AT-Informer model > Informer model.
e. By comparing the four evaluation criteria of MAE, MSE, RMSE and MAPE in Table 7, it is observable that in contrast to the other three models, except that the Informer-AT model exhibits the lowest MAPE in autumn or winter, the AT-Informer-AT model' prediction errors are the smallest across all four seasons, demonstrating that it possesses the best prediction performance and the greatest forecast precision among the four models.

Combined with the comparison outcomes of the aforementioned two markets, it is evident that our AT-Informer-AT model attains the utmost forecasting precision. The incorporation of AM layer into both the Informer model's encoder and decoder can effectively enhance the predictive efficacy as well as the model's generalized capability.

4.5.2. Horizontal comparison

To further reflect the efficacy of the proposed method, we select several common models including Informer, BiLSTM, BiGRU, TCN, XGBoost and SVM for comparison. Since previous studies have shown that the TCN model outperforms CNN [24], the BiLSTM model exhibits superior forecasting accuracy than the LSTM model [25], the BiGRU model demonstrates superior forecasting results in contrast to the GRU model [26], and the Informer model has superior prediction outcomes than Transformer model [27], our trials are no further include

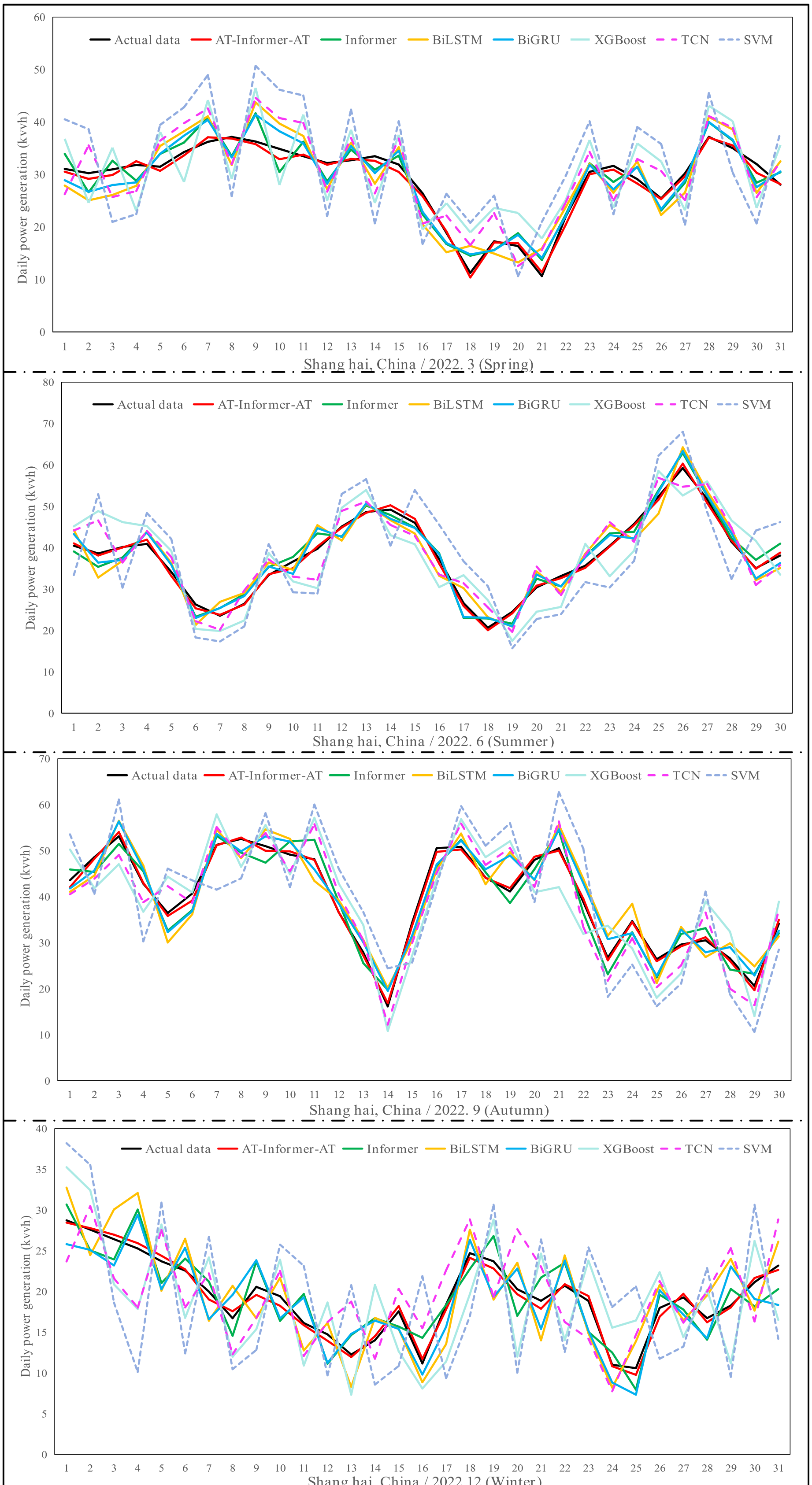


Fig. 20. Comparison between actual data and predicted results (Xuhui).

Fig. 21. The error comparison results (Xuhui).

comparisons with models like CNN, LSTM, GRU, and Transformer. This research is also carried out in two places: DKASC area and Xuhui District. The model forecasting outcomes for these two cases are as listed:

(1) DKASC area

The contrast among the true data and predicted outcomes from various models is illustrated in Fig. 18, while the forecasting error contrast is shown in Fig. 19. Here is the detailed analysis:

a. From Fig. 18, it is evident that in contrast to the different models' fitting degree as well as real data in four seasons, the put forward AT-Informer-AT model exhibits the most superior fitting effect and highest prediction accuracy between the seven models. Conversely, the forecasting outcomes of SVM, XGBoost, TCN models exhibit a subpar forecasting accuracy and an inadequate goodness of fit with real data.
b. By comparing the four evaluation criteria of MSE, MAE, RMSE and MAPE across four seasons in Fig. 19, it is evident that in contrast to the other models, the AT-Informer-AT model presents the smallest forecasting error, indicating it exhibits the best forecasting performance as well as prediction precision. In addition, the prediction errors of all models across four seasons are ranked as follows: 1) MSE: autumn > winter > summer > spring; 2) MAE: autumn > winter > summer > spring; 3) RMSE: autumn > winter > summer > spring; 4) MAPE: spring > autumn > winter > summer.
c. From the monthly perspective, the DKASC area experiences a tropical desert climate characterized by minimal rainfall and abundant solar radiation year-round. PV data exhibits minimal fluctuation in June and September, indicating relatively stable conditions, while March and December demonstrate pronounced fluctuations, possibly influenced by seasonal variations in solar intensity and other climatic factors.

By integrating the aforementioned forecasting outcomes, we observe that the proposed method demonstrates better prediction accuracy as well as enhanced performance. To comprehensively showcase the effectiveness of the put forward prediction approach, we select the forecasting outcomes from Xuhui District as validation below.

(2) Xuhui District

The contrast among the true data and predicted outcomes from various models is illustrated in Fig. 20, while the forecasting error contrast is shown in Fig. 21. Here is the detailed analysis:

a. From Fig. 20, it is evident that in contrast to the different models' fitting degree as well as real data in four seasons, the proposed AT-Informer-AT model demonstrates the most accurate alignment with the real data and provides the best fit, both in the overall trend and in specific amplified regions. Conversely, the forecasting outcomes from the SVM, XGBoost, and TCN models exhibit poor alignment with the real data and their predictive performance is subpar. This is consistent with the forecast trend of DKASC area.
b. Comparing the four evaluation criteria of MSE, MAE, RMSE and MAPE across four seasons in Fig. 21, it is evident that, in all error evaluation criteria, the AT-Informer-AT model exhibits the lowest prediction error in contrast to other six models, indicating it exhibits the best predition performance as well as prediction precision. Furthermore, the prediction errors of all models across four seasons are ranked as follows: 1) MSE: summer > spring > autumn > winter; 2) MAE: summer > spring > autumn > winter; 3) RMSE: summer > spring > autumn > winter; 4) MAPE: summer > winter > autumn > spring.
c. From a monthly perspective, in contrast to March as well as December, the fluctuation range and PV generation in June and September are larger. This observation aligns with the

climatic traits of Shanghai, where the flood season is focused between May and September, coinciding with stronger solar radiation during this period.

In accordance with the above analysis, the proposed method demonstrates the closest to overall trend of real data, with the most effective fitting. The forecasting outcomes also align with the climate change patterns of two cities. While the forecasting error is not significantly distirict with other models, it has the smallest value in most months as well as evaluation criteria. Based on the fitting degree and forecasting error with real data, the proposed method shows the best predictive efficacy, which proves beneficial in improving the precision of PVGF.

5. Conclusion

To enhance the precision of PV generation and decrease the risk of volatility, this paper proposes a novel Attention-Informer-Attention (AT-Informer-AT) method based on Informer model integrating Attention Mechanism. In data processing, this paper utilizes LOWESS to alleviate the volatility range of daily PV as well as boost data steadiness. Additionally, FE technique is employed to simplify important features and elevate the quality of input data. When it comes to model prediction, the Informer model serves as the foundation, with the integration of an AM layer into both the encoder and decoder modules. This augmentation aims to bolster prediction performance and enhance the model's generalization ability. The case analysis results from two different areas illustrate that the proposed forecasting approach notably enhances the prediction performance of PVGF, showing excellent forecasting accuracy and stability. The primary findings are outlined below:

(1) Under the premise of the same processing of PV data, in contrast to the basic Informer model, the AT-Informer model as well as the Informer-AT model, the proposed AT-Informer-AT model attains the utmost forecasting precision and the lowest forecasting errors across the most of evaluation criteria in four seasons.
(2) Compared with the single addition of AM layer, the incorporation of AM layer into the Informer model's encoder and decoder can more effectively enhance the model's predictive efficacy and generalizability.
(3) Under the same circumstances of dataset, the proposed AT-Informer-AT model exhibits significantly inferior forecasting error as well as superior forecasting accuracy than other six prediction models.
(4) In terms of the fitting degree with real data, the proposed method which integrates LOWESS smoothing and feature engineering technique demonstrates the most accurate alignment with the real data and the best fit, both in overall trend as well as localized amplification regions, indicating the highest forecasting performance than others.
(5) From the monthly perspective, the PV generation and its fluctuation range in different regions are closely related to the climatic characteristics of those regions. During months with strong solar radiation and climate variability, PV generation tends to be high, leading to a broader fluctuation range, alongside larger errors for the corresponding season.

Though the forecasting approach enhances the prediction precision of PVGF, there also exist some issues of concern. For example, how can we find a suitable algorithm to automatically adjust our parameters? How to comprehensively compare the proposed model with other Transformer variants or state-of-the-art PVGF methods? Furthermore, how to evaluate different feature engineering strategies to systematically investigate their influence on model performance? These will be the focus of our future research.

CRediT authorship contribution statement

Mingming Leng: Writing – review & editing. **dai yeming:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Tao Ren:** Software, Resources. **Weijie Yu:** Writing – original draft, Software, Resources, Methodology, Formal analysis, Conceptualization.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests. Yeming Dai reports financial support was provided by National Natural Science Foundation of China. Yeming Dai reports financial support was provided by Humanities and Social Science Fund of Ministry of Education of the People's Republic of China. Yeming Dai reports financial support was provided by Shandong Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the supports of National Natural Science Foundation of China (No.72371139), the Humanities and Social Science Fund of Ministry of Education of China (No.20YJA630009), Shandong Natural Science Foundation of China (No.ZR2022MG002).

Data availability

Data will be made available on request.

References

[1] G.Q. Lin, L.L. Li, M.L. Tseng, et al., An improved moth-flame optimization algorithm for support vector machine prediction of photovoltaic power generation, J. Clean. Prod. 253 (2020) 119966.
[2] U.G.K. Mulleriyawage, W.X. Shen, Optimally sizing of battery energy storage capacity by operational optimization of residential photovoltaic-Battery systems: an Australian household case study, Renew. Energy 160 (2020) 852–864.
[3] K.K. Tong, S.Y. Sun, Multi-dimensional decoupling analysis in the context of energy use: dynamic well-being, resource, and impact decoupling relationships in China, Appl. Energy 359 (2024) 122726.
[4] J. Li, Z. Fang, Y. Tan, et al., Application prospect analysis of electrochemical energy storage technology in county-wide rooftop photovoltaic system, Acta Energ. Sol. Sin. 43 (4) (2022) 1–12.
[5] Almaghrabi, S. Rana, M. Hamilton, et al., Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks, Inf. Fusion 104 (2024) 102180.
[6] Z.D. Zhou, Y.M. Dai, M.M. Leng, A photovoltaic power forecasting framework based on Attention mechanism and parallel prediction architecture, Appl. Energy 391 (2025) 125869.
[7] S. Qiao, L. Wang, P. Zhang, et al., Short-term load forecasting by long and short-term temporal networks with attention based on model decomposition, TPower Syst. Technol. 46 (10) (2022) 3940–3951.
[8] S. Ding, R. Li, Z. Tao, A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting, Energy Convers. Manag. 227 (2021) 113644.
[9] R. Ahmed, V. Sreeram, Y. Mishra, et al., A review and evaluation of the state-of-the-art in photovoltaic solar power forecasting: techniques and optimization, Renew. Sustain. Energy Rev. 124 (2020) 109792.
[10] Y. Gong, Z. Lu, Y. Qiao, et al., An overview of photovoltaic energy system output forecasting technology, Autom. Electr. Syst. 40 (4) (2016) 140–151.
[11] X.Z. Wang, C.W. Li, C.Q. Yi, et al., EcoForecast: An interpretable data-driven approach for short-term macroeconomic forecasting using N-BEATS neural network, Eng. Appl. Artif. Intell. 114 (2022) 105072.
[12] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting[J], Energy 283 (2021) 116239.
[13] Q. Liu, Y.L. Li, H. Jiang, et al., Short-term photovoltaic power forecasting based on multiple mode decomposition and parallel bidirectional long short term combined with convolutional neural networks, Energy 286 (2024) 129580.
[14] A. Gasparin, S. Lukovic, C. Alippi, Deep learning for time series forecasting: the electric load case, CAAI Trans. Intell. Technol. 7 (1) (2022) 1–25.
[15] Z. Wang, H. Liu, S. Wu, et al., Explainable time-varying directional representations for photovoltaic power generation forecasting, J. Clean. Prod. 468 (2024) 143056.

[16] Z. Zhao, Y. Zhang, Y. Yang, et al., Load forecasting via grey model-least squares support vector machine model and spatial-temporal distribution of electric consumption intensity, Energy 255 (2022) 124468.
[17] J. Xu, J. He, J. Gu, et al., Financial time series prediction based on XGBoost and generative adversarial networks, Int. J. Circuits, Syst. Signal Process. 16 (2022) 637–645.
[18] W. Guo, L. Che, M. Shahidehpour, et al., Machine-learning based methods in short-term load forecasting, Electr. J. 34 (1) (2021) 106884.
[19] X. Ren, F. Zhang, H. Zhu, et al., Quad-kernel deep convolutional neural network for intra-hour photovoltaic power forecasting, Appl. Energy 323 (2022) 119682.
[20] W.L. Zeng, Q.Y. Lin, B.Y. Zhu, et al., Modeling vehicle U-turning behavior near intersections: A deep learning approach based on TCN and multi-head attention, Expert Syst. Appl. 249 (2024) 123674.
[21] T. Peng, Chu Zhang, J.Z. Zhou, et al., An integrated framework of Bi-directional long-short term memory (BiLSTM) based on sine cosine algorithm for hourly solar radiation forecasting, Energy 221 (2021) 119887.
[22] Y. He, Y. Wang, Short-term wind power prediction based on EEMD-LASSO-QRNN model, Appl. Soft Comput. 105 (2021) 107288.
[23] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, Neural Inf. Process. Syst. (2017) 5998–6008.
[24] T.R. Zhang, Y.P. Huang, H. Liao, et al., A hybrid electric vehicle load classification and forecasting approach based on GBDT algorithm and temporal convolutional network, Appl. Energy 351 (2023) 121768.
[25] V.M. Joshi, R.B. Ghongade, A.M. Joshi, et al., Deep BiLSTM neural network model for emotion detection using cross-dataset approach, Biomed. Signal Process. Control 73 (2022) 103407.
[26] X. Dong, H. Zhao, S. Zhao, et al., ULTRA-SHORT-TERM forecasting method of photovoltaic power based on som clustering, secondary decomposition and BiGRU, Acta Energ. Olaris Sin. 43 (11) (2022) 85–93.
[27] H. Zhou, S. Zhang, J. Peng, et al., Informer: beyond efficient transformer for long sequence time-series forecasting, Proc. AAAI Conf. Artif. Intell. 35 (12) (2021) 11106–11115.
[28] Z.Y. Shi, J. Li, Z.Y. Jiang, et al., WGformer: a weibull-gaussian informer based model for wind speed prediction, Eng. Appl. Artif. Intell. 131 (2024) 107891.
[29] M.J. Gong, Y. Zhao, J.W. Sun, et al., Load forecasting of district heating system based on Informer, Energy 253 (2022) 124179.
[30] X.M. Xiu, L.Y. Guan, Z.Y. Wang, et al., A double-layer forecasting model for PV power forecasting based on GRU-Informer-SVR and Blending ensemble learning framework, Appl. Soft Comput. 172 (2025) 112768.
[31] Y.S. Cao, G. Liu, D.H. Luo, et al., Multi-timescale photovoltaic power forecasting using an improved Stacking ensemble algorithm based LSTM-Informer model, Energy 283 (2023) 128669.
[32] S. Cui, S. Lyu, Y. Ma, et al., Improved Informer PV power short-term prediction model based on weather typing and AHA-VMD-MPE, Energy 307 (2024) 132766.
[33] S.D. Du, T.R. Li, Y. Yang, et al., Multivariate time series forecasting via attention-based encoder-decoder framework, Neurocomputing 388 (2020) 269–279.
[34] Z.J. Ma, G. Mei, A hybrid attention-based deep learning approach for wind power prediction, Appl. Energy 323 (2022) 119608.
[35] Z.A. Khan, T. Hussain, S.W. Baik, Dual stream network with attention mechanism for photovoltaic power forecasting, Appl. Energy 338 (2023) 120916.
[36] L. Wang, M. Mao, J. Xie, et al., Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model, Energy 262 (2023) 125592.
[37] Y. Wang, H. Wang, D. Srinivasan, et al., Robust functional regression for wind speed forecasting based on Sparse Bayesian learning, Renew. Energy 132 (2019) 43–60.
[38] K. Chen, S. Laghrouche, A. Djerdir, Health state prognostic of fuel cell based on wavelet neural network and cuckoo search algorithm, ISA Trans. 113 (2021) 175–184.
[39] PirhooshyaranM, K. Scheinberg, L.V. Snyder, Feature engineering and forecasting via derivative-free optimization and ensemble of sequence-to-sequence networks with applications in renewable energy, Energy 196 (2020) 117136.
[40] Y.M. Dai, Y.X. Wang, M.M. Leng, et al., LOWESS smoothing and Random Forest based GRU model: A short-term photovoltaic power generation forecasting method, Energy 256 (2022) 124661.
[41] B. Yu, J. Li, C. Liu, et al., A novel short-term electrical load forecasting framework with intelligent feature engineering, Appl. Energy 327 (2022) 120089.
[42] W. Huo, W. Li, Z. Zhang, et al., Performance prediction of proton-exchange membrane fuel cell based on convolutional neural network and Random Forest feature selection, Energy Convers. Manag. 243 (2021) 114367.
[43] S.C. De Lima Nogueira, S.H. Och, L.M. Moura, et al., Prediction of the NOx and CO2 emissions from an experimental dual fuel engine using optimized random forest combined with feature engineering, Energy 280 (2023) 128066.
[44] Y.M. Dai, W.J. Yu, M.M. Leng, A hybrid ensemble optimized BiGRU method for short-term photovoltaic generation forecasting, Energy 299 (2024) 131458.
[45] R.G. Da Silva, M.H.D. Ribeiro, S.R. Moreno, et al., A novel decomposition-ensemble learning framework for multi-step ahead wind energy forecasting, Energy 216 (2021) 119174.
[46] ⟨https://www.lvsedianli.com/perHome.html⟩.
[47] ⟨https://dkasolarcentre.com.au/download/notes-on-the-data⟩.
[48] ⟨https://power.larc.nasa.gov/data-access-viewer/⟩.