Renewable Energy 258 (2026) 125012

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term photovoltaic forecasting: A parallel TimesNet and AT-Informer-AT method

Weijie Yu [a,b], Yeming Dai [a,*], Wenjie Wang [b,**], Tao Ren [c], Mingming Leng [d]

[a] School of Business, Qingdao University, Qingdao, 200071, China
[b] Glorious Sun School of Business and Management, Donghua University, Shanghai, 200051, China
[c] College of Computer Science and Technology, Qingdao University, Qingdao, 200071, China
[d] Faculty of Business, Lingnan University, Hong Kong

ARTICLE INFO

Keywords:
Photovoltaic
Locally weighted scatterplot smoothing
Feature engineering
AT-Informer-AT
TimesNet

ABSTRACT

As the integration of renewable energy accelerates, high accuracy Photovoltaic Power Generation Forecasting (PVGF) has become a key enabler for maintaining grid resilience and planning energy dispatch efficiently in distributed power network. To further improve forecasting performance and stability, we have developed a short-term PVGF method with a parallel architecture. Initially, Locally Weighted Scatterplot Smoothing (LOWESS) is applied to reduce data noise and stabilize the input sequences. Moreover, Feature Engineering (FE) is utilized to identify the most relevant input variables. Thirdly, a parallel model named 'TNet-AIA' is designed, which incorporates a parallel structure combining the strengths of TimesNet and Attention-Informer-Attention (AT-Informer-AT) models. Specifically, the TimesNet model is employed to capture multi-scale temporal patterns in the input sequences, while the AT-Informer-AT model successfully learns both long-term correlations and short-term local variations. Case studies are conducted on two representative Photovoltaic Power (PV) located in DKASC area, Alice Springs, Australia, and Xuhui District, Shanghai, China. Experimental findings indicate that the presented approach significantly improves the predictive performance and stability, achieving a notable 16.88 % improvement in forecasting accuracy.

Nomenclature

AM	Attention Mechanism	MAPE	Mean Absolute Percentage Error
AT-Informer-AT	Attention-Informer-Attention	ML	Machine Learning
BiGRU	Bi-directional Gate Recurrent Unit	MSE	Mean Square Error
BiLSTM	Bi-directional Long Short-Term Memory	NN	Neural Network
CNN	Convolutional Neural Network	PCA	Principal Component Analysis
DKASC	Desert Knowledge Australia Solar Centre	PC1	first Principal Component
FC	Feature Construction	PV	Photovoltaic Power
FE	Feature Engineering	PVGF	Photovoltaic Power Generation Forecasting

(*continued on next column*)

(*continued*)

FS	Feature Selection	Q1	First Quartile
GRU	Gate Recurrent Unit	Q3	Third Quartile
IEA	International Energy Agency	RF	Random Forests
IQR	Interquartile Range	RMSE	Root Mean Square Error
LOWESS	Locally Weighted Scatterplot Smoothing	SVM	Support Vector Machine
LSTM	Long Short-Term Memory	TCN	Temporal Convolutional Network
MAE	Mean Absolute Error	XGBoost	eXtreme Gradient Boosting

1. Introduction

PV, as a green and sustainable energy source, has seen rapid global

* Corresponding author.
** Corresponding author.
E-mail addresses: weijieyu212@163.com (W. Yu), yemingdai@163.com (Y. Dai), wenjiew@dhu.edu.cn (W. Wang), rentao@qdu.edu.cn (T. Ren), mmleng@ln.edu.hk (M. Leng).

https://doi.org/10.1016/j.renene.2025.125012
Received 15 August 2025; Received in revised form 1 December 2025; Accepted 12 December 2025
Available online 15 December 2025
0960-1481/

adoption in recent years. The Electricity 2025 report released by IEA, PV is forecast to meet roughly half of the growth in global electricity demand through to 2027 [1]. However, the inherent variability and intermittency of PV output, largely influenced by weather conditions, pose significant challenges for power system operation and grid stability. Therefore, accurate PVGF has become essential for ensuring dependable power system operation and enabling the large-scale integration of solar energy into modern electricity networks [2].

Many studies have indicated that the performance of PVGF is strongly influenced by both the forecasting interval and the chosen prediction method [3]. Based on the length of prediction horizon, PVGF is typically categorized into ultra-short-term [4], short-term [5], and medium to long-term [6] categories. Specifically, ultra-short-term prediction refers to a few seconds to several minutes, short-term spans several minutes to two days, while medium to long-term extends from several days to multiple years. Selecting an appropriate forecasting horizon is crucial to achieving accurate PVGF [7]. In particular, short-term prediction serves as a key component in supporting real-time electricity scheduling and addressing immediate demand from downstream users. However, its accuracy is often hindered by the intrinsic variability of solar power and the influence of rapidly changing meteorological conditions. To overcome these challenges, accurate PVGF requires not only high-quality data preprocessing to reduce input uncertainty, but also the design of robust models capable of capturing complex temporal patterns and learning effectively from noisy or incomplete information.

In data processing, PV data often exhibits significant variability and irregularity caused by environmental influences, including solar irradiance and humidity. Therefore, stabilizing the input data is crucial in PVGF, particularly when large fluctuations are present [8]. To deal with the challenges of data instability and volatility, data processing approaches such as smoothing denoising [9] and FE [10] have proven effective in enhancing data quality and improving subsequent model input reliability [11]. Our earlier study [12] demonstrated that LOWESS is particularly effective for PV data, significantly mitigating seasonal fluctuations and enhancing the forecasting performance. In addition, FE contributes to improving model accuracy and data interpretability by transforming and generating new features from initial data. Generally positioned during data processing, FE involves two key components [13]: FC and FS, the combination of which helps to make the data more informative and conducive to learning [14].

In model prediction, PVGF methods are broadly categorized as either physical [15] or data-driven [16]. Physical approaches tend to be complex and highly sensitive to input uncertainties, making them inadequate for short-term PVGF [17]. In contrast, data-driven approaches, including statistical models [18] and ML techniques [19], have gained increasing attention due to their adaptability and predictive accuracy. Statistical models based on time series analysis [20] can capture temporal patterns, but often neglect exogenous factors, reducing forecasting accuracy. As a subset of artificial intelligence forecasting, ML approaches have shown strong performance in PVGF, with commonly used models including SVM [21], XGBoost [22], and various forms of NN [23]. Among them, NN is particularly favored for its strength to process large data as well as capture nonlinear interactions. Classical NN employed in PVGF include TCN [24], BiLSTM [25], BiGRU [26], and Transformer [27]. Transformer, though effective in natural language processing, faces limitations in time series forecasting due to high time and memory complexity. To tackle the issue, Zhou et al. [28] proposed Informer model, which improves efficiency and supports long-sequence forecasting, making it increasingly popular in time series applications. Its scalability and effectiveness have since drawn increasing interest in time series applications, such as wind power prediction [29], district heating systems [30], and PVGF [31]. These studies highlight Informer's strong predictive capability and motivate its application in energy forecasting. Similarly, AM based models have gained increasing traction in time series prediction, as it effectively captures complex time-dependent patterns. To illustrate, Du et al. [32] introduced a temporal AM encoder-decoder framework tailored for multivariate time series prediction, demonstrating strong predictive performance. In the context of PVGF, Khan et al. [33] developed a parallel architecture embedding AM to enhance forecasting accuracy. Therefore, our previous study proposed an innovative AT-Informer-AT framework, which extends the Informer architecture by embedding AM to improve the accuracy of PVGF and mitigate volatility risks [34]. However, it still faces limitations in efficiency and stability. Here, stability is defined as the model's ability to maintain consistent forecasting performance across different time periods and datasets, reflecting robustness under varying conditions. In contrast, TimesNet model is a recently proposed deep learning architecture designed for time series forecasting [35]. It can effectively learn enduring temporal relationships and complex structures in long sequences and high-dimensional data, enhancing forecasting efficiency and stability [36].

With the aim of enhancing the accuracy and stability of PVGF, based on the previous study, we have proposed a novel parallel architecture prediction method, integrating the TimesNet and AT-Informer-AT models. LOWESS is initially applied to process the PV data, and FE is then applied to optimize the data, ensuring high-quality inputs. Moreover, a parallel model is designed, which incorporates a parallel structure combining the strengths of TimesNet and AT-Informer-AT models. Finally, through conducting comparative analyses on two regional datasets covering four seasons, we have verified that the proposed approach leads to significant improvements in PVGF, particularly in terms of forecasting precision and consistency. The core contributions of this study are as follows:

(1) A novel parallel architecture, TNet-AIA prediction method, is introduced to further enhance the precision and stability of AT-Informer-AT model.
(2) The parallel prediction framework incorporates the TimesNet and AT-Informer-AT models.
(3) The proposed method's effectiveness is validated on seasonal PV datasets from two diverse regions, through comparisons with other models.

This study is divided into the following sections: Section 2 introduces the core principles of LOWESS, FE, AT-Informer-AT and TimesNet model. Section 3 details the proposed Parallel prediction method and the detailed process. Section 4 provides an in-depth experimental comparison to evaluate the effectiveness of the proposed method. Section 5 concludes the paper with a summary of our findings.

2. Theoretical basis

This section presents the essential approaches utilized for data processing and prediction. The data processing pipeline incorporates LOWESS and Feature Engineering, while the forecasting task is conducted using AT-Informer-AT and TimesNet as baseline models.

2.1. LOWESS

LOWESS is a non-parametric regression approach commonly used for smoothing and prediction. For each data point x, a localized linear regression is performed on neighboring observations, with weights assigned based on their distance from x. These weights are determined by a kernel function, typically a cubic form as defined in Eq. (1). This process is repeated for all n data points, resulting in a set of n locally weighted regression curves [37].

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

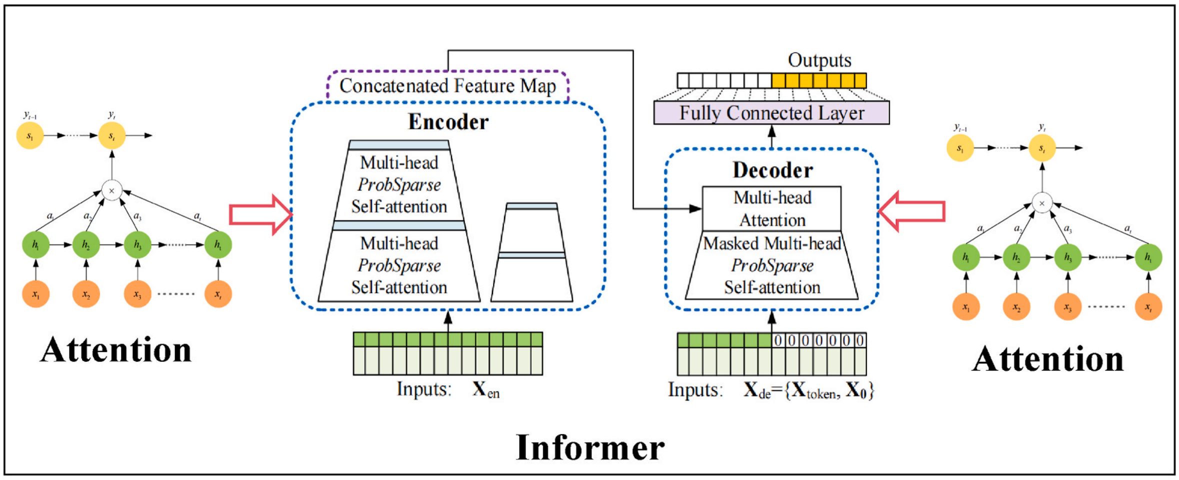


Fig. 1. The framework of AT-Informer-AT model.

2.2. Feature Engineering

FE plays a vital role in machine learning and data extraction by identifying and extracting informative features relevant to the target variable, thereby enhancing model performance and predictive accuracy [10]. Generally, FE comprises two core modules: FC and FS [13].

FC aims to enhance predictive power by generating new features through various techniques, among which PCA is commonly adopted in various studies [38]. To reduce the dimensionality of data, PCA linearly converts high dimensions data to extract principal components by computing the eigenvectors of the input covariance matrix. Following FC, FS removes redundant or irrelevant features to retain informative subsets, thereby improving model performance and forecasting accuracy [13]. Among FS techniques, RF is frequently employed due to its ability to evaluate variable importance in high-dimensional data. By constructing various decision trees via sampling randomly and selecting features, RF determines feature relevance based on changes in out-of-bag errors when specific variables are randomly permuted [39]. The main equation is as listed:

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

where, VI represents the significance of data, $error1$ and $error2$ represent two kinds of out-of-bag errors in RF, while N denotes the decision trees' number.

2.3. AT-Informer-AT model

The AT-Informer-AT model adds AM layers to the encoder and decoder of the Informer model to make it more focused on time steps that have an important impact on the prediction target, and more accurately capture the information related to the practice sequence data, enhancing the model's prediction and generalization performance [34]. The structure of AT-Informer-AT model is shown in Fig. 1. The core equation is as equation (3).

$$Attention(Q, K, V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3}$$

where Q, K, and V denote the matrixes of Query, Key, and Value, while d_k represents the key's dimension.

2.4. TimesNet model

The TimesNet model adopts a modular architecture and consists of a series of TimesBlocks stacked via residual pathways, which can decompose complex temporal variations into distinct periods [35]. It maps the original one dimension time series into a two dimension representation, enabling joint modeling of variations both within and across cycles. By capturing the correlation of these cycles, the prediction accuracy can be effectively enhanced [36]. The framework TimesNet model is shown in Fig. 2.

$$X_{2D}^i = \mathrm{Reshape}_{p_i, f_i}(Padding(X_{1D})), i \in \{1, \ldots, k\} \tag{4}$$

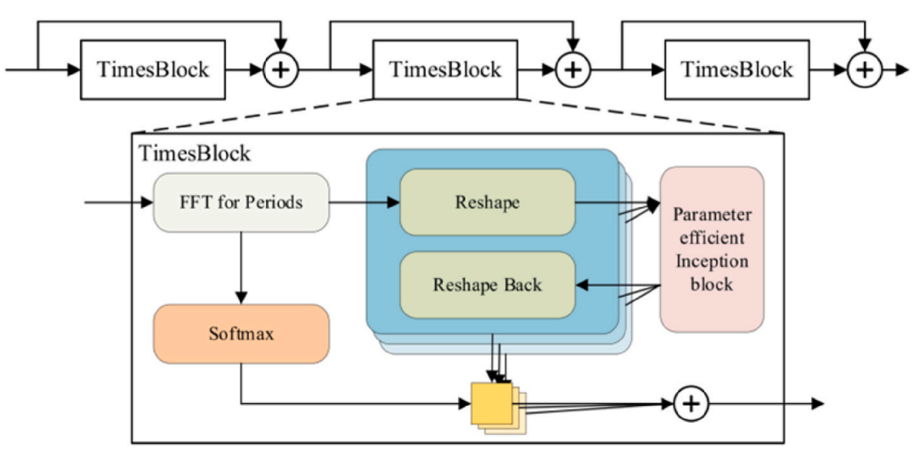


Fig. 2. The framework of TimesNet model.

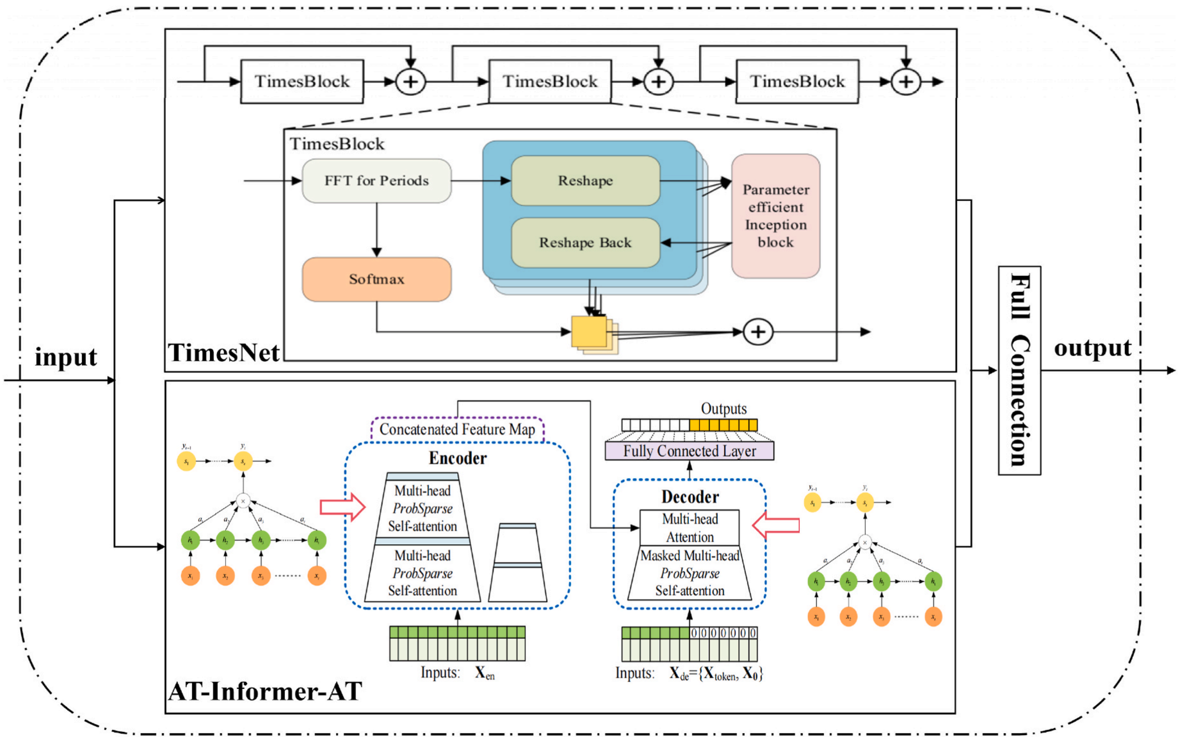


Fig. 3. The framework of TNet-AIA model.

$$\hat{X}^{i}_{2D} = Inception\left(X^{i}_{2D}\right), i \in \{1, \ldots, k\} \tag{5}$$

$$\hat{X}^{i}_{1D} = Trim\left(Reshape\left(\hat{X}^{i}_{2D}\right)\right), i \in \{1, \ldots, k\} \tag{6}$$

where, *Padding*() is a padding function that adds zeros at the sequence's end X_{1D} to make its length divisible. *Inception* denotes the TimesBlock block used to capture the tensor's feature representation after obtaining two-dimensional tensor, which captures temporal features panning both intra-period and inter-period dynamics in the series. *Trim*() denotes the zero deleted in formula (5).

3. The parallel prediction method

3.1. The AT-Informer-AT parallel TimesNet prediction model

Inspired by the research of Tian et al. [40], this paper combines the TimesNet model and the AT-Informer-AT model to design a parallel structure, which is denoted as ' TNet-AIA ' parallel model. Its structural framework is demonstrated in Fig. 3.

The forecasting procedure of the TNet-AIA parallel model is to input the processed photovoltaic power data into the TimesNet model and the AT-Informer-AT model respectively for parallel prediction, and connect the output of the two models with the help of a fully connected unit, and finally generate the forecasting results. Specifically, the TimesNet model focuses on processing sequence data with multi-scale temporal characteristics, while the AT-Informer-AT model combines the AM layer to improve the perception of long-term dependence and local fluctuations. The two models exert their respective advantages, predict separately, and then fuse their outputs through the full connection layer. The role of the fully connected unit is to weight and integrate the outputs from the two sub-models and generate the final PV power supply prediction results.

3.2. The TNet-AIA parallel forecasting method

Fig. 4 presents the overall workflow of the TNet-AIA parallel forecasting method, corresponding to the predictive framework introduced in Section 3.1. The forecasting process is structured into the following three primary stages:

Step 1. Data processing. To ensure high-quality inputs for model training, a comprehensive data processing pipeline is employed. Initially, the initial PV data undergo cleaning, down-sampling, and min-max normalization to address missing values, outliers, and erroneous entries, effectively minimizing noise and enhancing data reliability. Subsequently, the LOWESS is applied to reduce daily fluctuations in PV output and improve temporal consistency. In addition, FE is employed to identify and refine relevant input variables. Specifically, PCA is utilized to compress the smoothed PV data by removing features with low or negligible correlation. The remaining features are then refined by RF, which ranks and selects the crucial features to boost data quality.
Step 2. Model training. The TNet-AIA forecasting model is implemented. The processed PV data will be input into the TNet-AIA parallel model for parallel prediction, and the final prediction result will be output through the full connection layer connection.
Step 3. Forecast Evaluation and Comparison. Evaluation of the forecasts is conducted through MSE, RMSE, MAPE, and MAE. Subsequently, the box plot is used to evaluate the stability of the models, so as to visually present the comparative performance across models, and verify the stability and superiority of the proposed method in dealing with different data sets.

4. Case study

4.1. Data preparation

To ensure data comprehensiveness and representativeness, we utilize datasets from two different PV stations: DKASC area, Alice Springs,

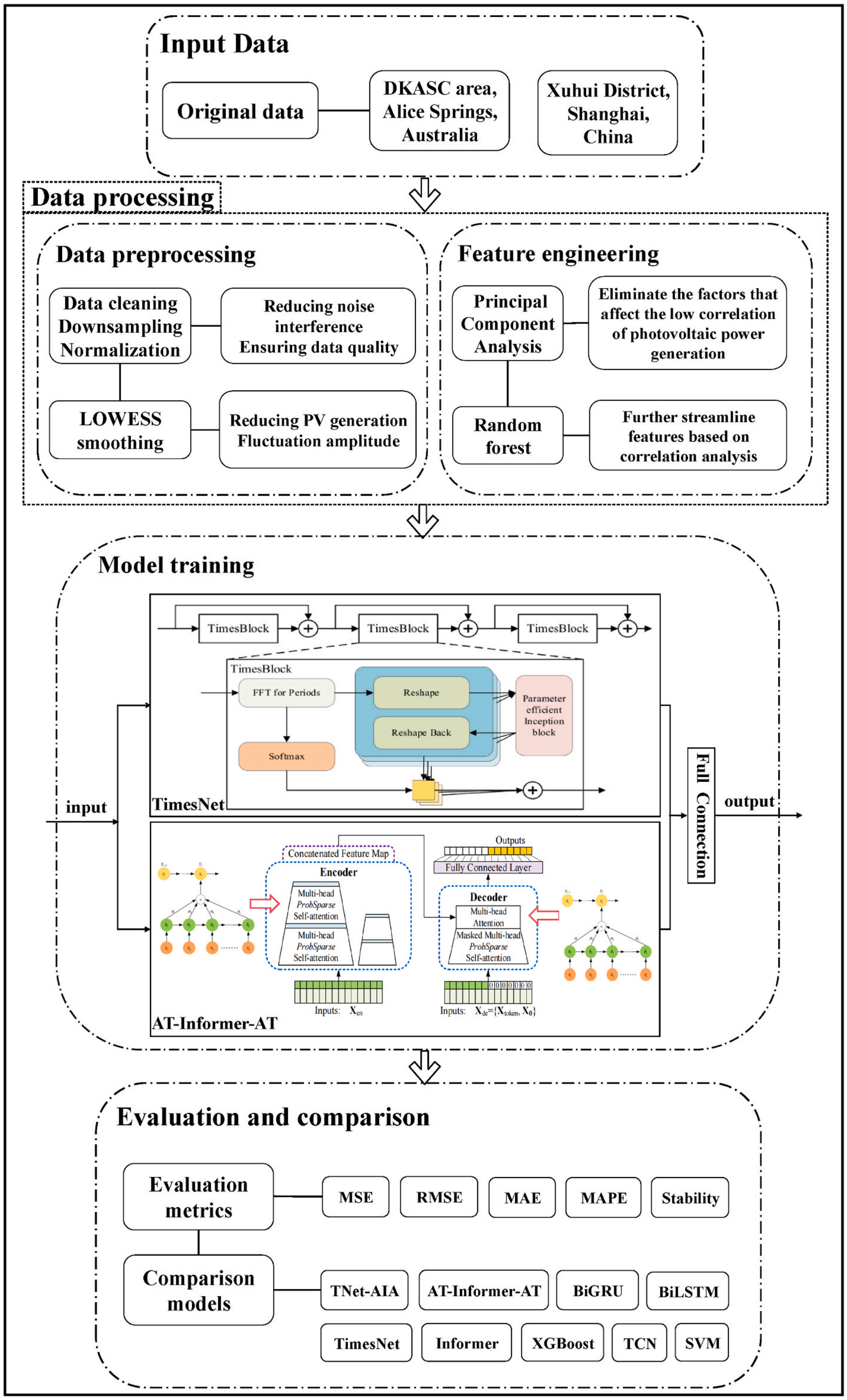
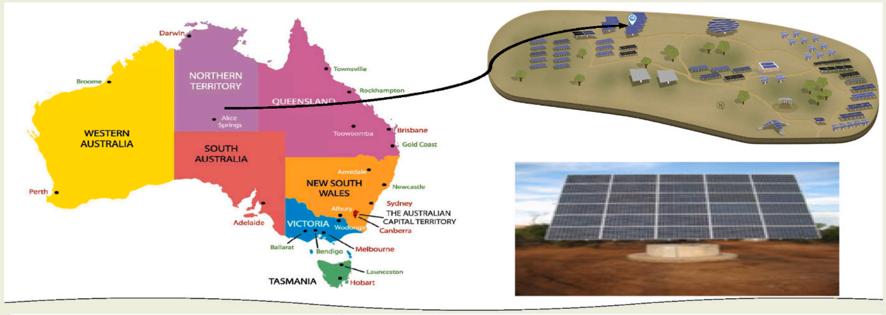


Fig. 4. The overall process of proposed method.

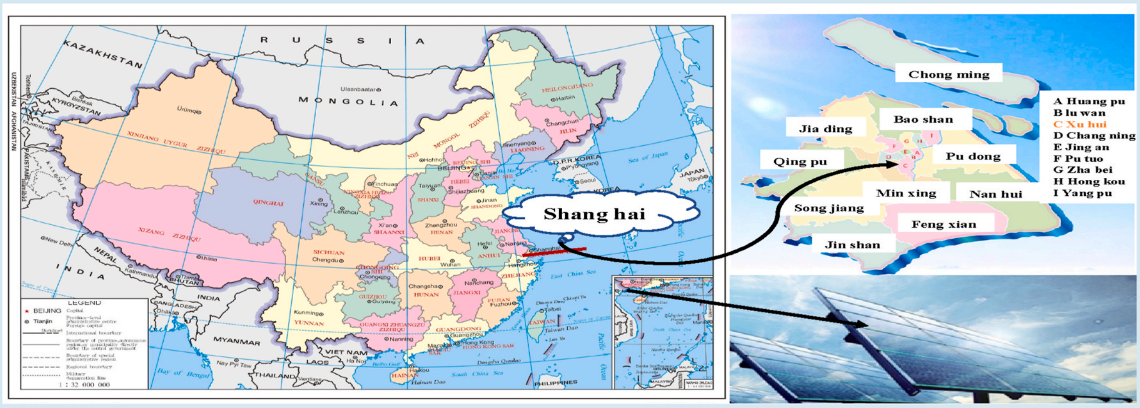
Australia [41] and Xuhui District, Shanghai, China [42]. These two locations were deliberately chosen due to their favorable climatic conditions, abundant solar radiation, and comparable latitudinal positions. Notably, DKASC lies in the Southern and Western Hemispheres, whereas Xuhui is positioned in the Northern and Eastern Hemispheres. This geographic diversity enhances the datasets' generalizability and representativeness.

To rigorously assess the performance of our proposed method, we have employed PV data spanning from 2015 to 2022 for both stations. The data is partitioned into training, validation, and test sets using a 7:1:2 ratio. Figs. 5 and 6 provide detailed illustrations of the PV sites' geographical locations, plant configurations, and the specific data splitting strategy. All data were collected on a daily basis. To prevent data leakage and ensure causal preprocessing, all data transformations (including LOWESS smoothing, normalization, and feature selection) are fitted exclusively on the training set and then applied unchanged to the validation and test sets.



Array	23.4KW
Panel rating	1.96W
Number of panels	4×30
Panel type	Trina TSM-195DCO1A
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.

4.2. Data processing

To ensure the accuracy and reliability of PVGF, comprehensive data preprocessing steps are essential. Building on our prior finding [12], which demonstrated the effectiveness of LOWESS in reducing data anomalies and enhancing the stability of PV time series, we have adopted a multi-step data preparation procedure. Initially, raw PV data undergo data cleaning, down-sampling, and min-max normalization to eliminate missing values, outliers, and erroneous entries, thereby mitigating noise and improving data quality. Subsequently, the LOWESS method is applied to further smooth the data, suppressing daily fluctuations and enhancing temporal consistency. We use datasets from

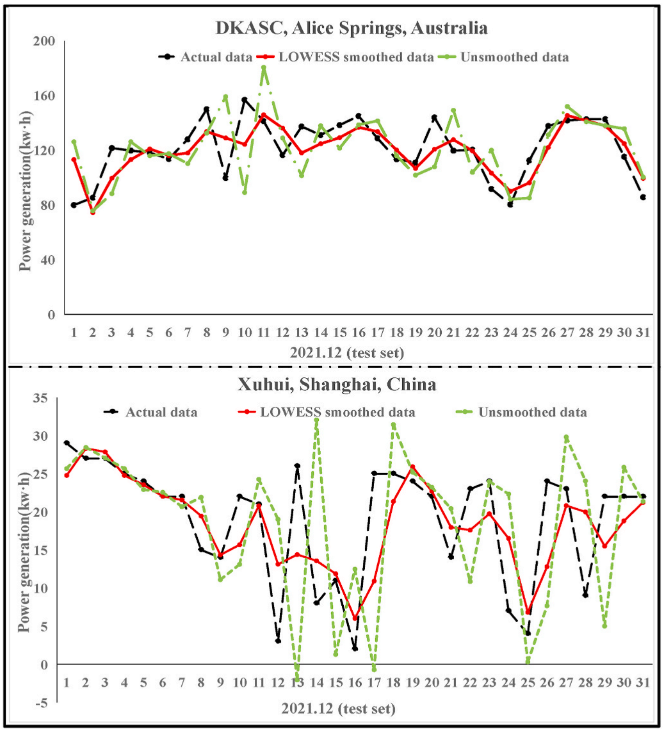


Fig. 7. Display of the smoothing impact.

DKASC and Xuhui, both collected in December 2021, to illustrate the smoothing effect. As shown in Fig. 7, the LOWESS-smoothed data exhibit tighter alignment with actual observations and reduced volatility compared to the unsmoothed series. The corresponding forecasting results confirm a notable improvement in prediction accuracy, underscoring the necessity of incorporating LOWESS into our preprocessing pipeline.

To further enhance predictive performance, we have implemented a FE process that identifies and refines relevant input variables. Drawing upon prior study [34], we have categorized the influencing factors into five main groups, and compile an initial feature set comprising 33 refined variables, as detailed in Table 1. Subsequently, PCA is employed to filter out features with low or no correlation, retaining only those whose absolute loading on the first principal component (PC1) exceeds 0.4. Specifically, "$|r|>0.4$" refers to the absolute PC1 loading. An original variable was kept for the subsequent RF analysis only if its PC1 loading exceeded this threshold. Moreover, RF is applied for importance ranking and selection, resulting in a final prioritized set of features. The results of this feature-engineering process are presented in Table 2, where A-E correspond to categories 1–5 in Table 1, respectively. Notably, the RF ranking consistently used the original variables rather than principal components. In the "PCA + RF", RF was trained on the subset of original variables that passed the PC1 loading filter; in the "No PCA + RF", it used all original variables. The random seed was fixed at 42 for reproducibility.

4.3. Parameter settings

The parameter settings play a critical role in determining the predictive performance of deep learning models. Proper configuration of model parameters directly affects convergence behavior, generalization ability, and overall accuracy. In this study, the AT-Informer-AT and TimesNet models were configured based on prior empirical findings and best practices. Specifically, the AT-Informer-AT model follows the configuration guidelines outlined in Ref. [40], while the TimesNet model adopts settings as suggested in Ref. [36]. The detailed parameter configurations for both models are presented in Tables 3 and 4, respectively.

Table 1
Category 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
Feature Engineering results.

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

4.4. Evaluation criteria

In this study, MAE, MSE, RMSE and MAPE are employed to evaluate the forecasting models, where y_i and y_i' in Eq. (7)–(10) denote the true and predicted values, respectively.

Table 3
Parameter settings of AT-Informer-AT 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 function	MSE

Table 4
Parameter settings of TimesNet.

Parameter	Parameter meaning	The parameter value
d_model	The dimension size of the vector in the model	32
d_ff	The dimension size of hidden layer in Feed-Forward network	32
enc_layers	Layers of encoder.	2
dec_layers	Layers of decoder.	1
n_heads	Number of heads	8
batch_size	Batch size	512
learning_rate	Learning rate	0.001
train_eopchs	Epochs	20
dropout	Dropout rate	0.1

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

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

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

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

Combining diverse evaluation metrics enables a comprehensive assessment of model performance. Among them, MAE measures the average absolute error, providing robustness against large deviations. MSE quantifies the squared differences, offering an intuitive and reliable measure of overall error. RMSE, sensitive to outliers, presents errors in the same units as the data, enhancing interpretability. MAPE expresses average relative error, making it suitable for comparing models across datasets with varying scales.

4.5. *Experiment results and comparative analysis*

To comprehensively assess the proposed model's performance, comparative analyses are conducted across multiple seasons. Specifically, data from the DKASC and Xuhui PV stations are employed, corresponding to March, June, September, and December of 2022. These months are selected to represent the four typical seasons: spring (March), summer (June), autumn (September), and winter (December). This seasonal segmentation enables a balanced evaluation under varying climatic and temporal conditions, as each month encapsulates the representative meteorological characteristics of its respective season. Despite the geographical placement of DKASC and Xuhui in opposite hemispheres, this selection remains appropriate, as it captures key seasonal variations in solar irradiance and weather patterns that influence photovoltaic system behavior. Consequently, the selected months provide a robust basis for analyzing seasonal effects on photovoltaic power output and facilitate meaningful inter-site comparisons.

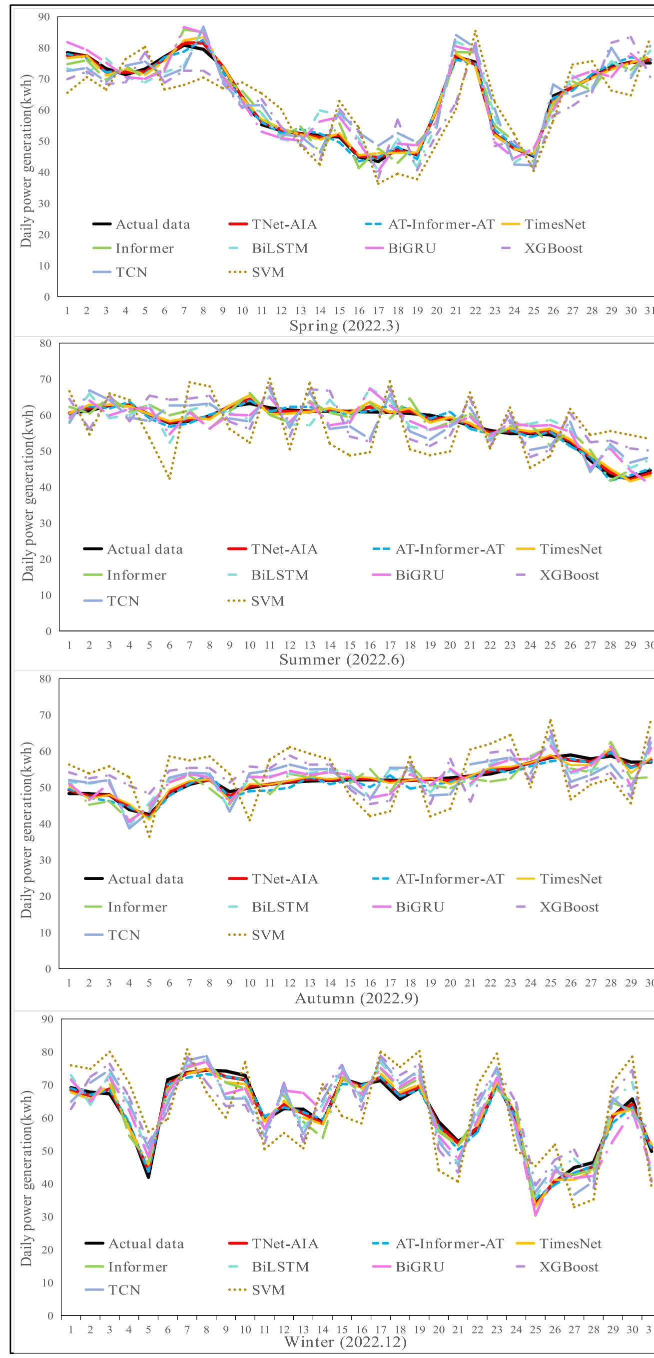


Fig. 8. Comparison between actual data and predicted results (DKASC).

In addition, to validate the proposed method's effectiveness, a series of widely used benchmark models are selected for comparison, including TimesNet, AT-Informer-AT, Informer, BiLSTM, BiGRU, TCN, XGBoost, and SVM. To ensure a fair comparison, all models are provided with the same time window as input, which includes the complete information of all time steps and all features. It is worth noting that the input window of all models refers to the historical observation sequence used for training and prediction, and its length matches the actual number of days in each experimental month. The data sampling frequency in this study is one

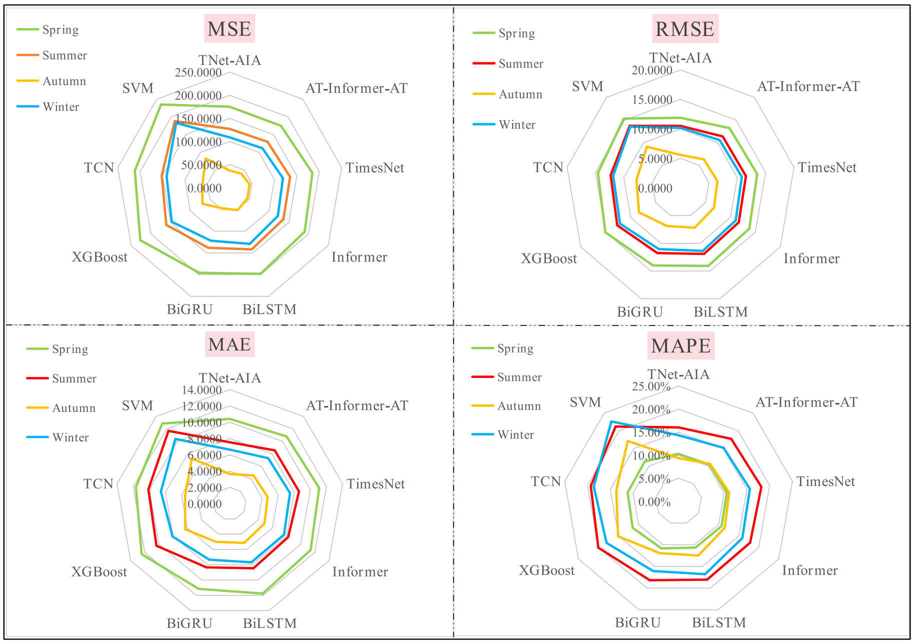


Fig. 9. Error comparison radar chart (DKASC).

day, meaning that all photovoltaic data were collected on a daily basis. The prediction horizon is set to one day, which corresponds to the short-term forecasting category in commonly adopted photovoltaic forecasting terminology. Given prior research indicates that TCN outperforms CNN [24], BiLSTM achieves higher prediction accuracy than LSTM [25], BiGRU exhibits improved results over GRU [26], and Informer surpasses the standard Transformer model [28]. Therefore, these models are not considered in our experimental comparisons. The findings from two experiments are as follows:

4.5.1. Comparative analysis of forecasting

(1) DKASC area, Alice Springs, Australia

A visual comparison of actual data versus predictions from multiple models is provided in Fig. 8, with Fig. 9 illustrating the differences in forecasting errors. The analysis is elaborated as follows:

A. As illustrated in Fig. 8, a comparison of the models' fitting performance across the four seasons reveals that the proposed TNet-AIA model achieves the most accurate forecasting and demonstrates the best overall fit to the observed data among the 10 models evaluated. On the contrary, the forecasting outcomes of SVM shows the lowest prediction accuracy and the worst fitting degree with actual data.

B. A comparative analysis of the four evaluation metrics across the four different seasons was conducted in radar chart Fig. 9, it can be seen that the proposed TNet-AIA model has the best in all evaluation indicators and has the lowest prediction error.

C. In the single model comparison, TimesNet has the best overall performance, Informer is slightly worse than TimesNet, but Informer is slightly better than TimesNet in the summer MAPE index; the overall performance of BiLSTM and BiGRU is close, and BiGRU is slightly better than BiLSTM. SVM performs worst on the whole, followed by XGBoost. Therefore, parallel TimesNet demonstrates its effectiveness in boosting the model's forecasting precision.

Based on the integrated forecasting results, the proposed method exhibits superior predictive accuracy. To further substantiate its effectiveness, we have presented the forecasting results from Xuhui District as a representative validation case.

(2) Xuhui District, Shanghai, China

A visual comparison of actual data versus predictions from multiple models is provided in Fig. 10, with Fig. 11 illustrating the differences in forecasting errors. The analysis is elaborated as follows:

A. It is evident from Fig. 10, comparing the fitting degree of different models with the actual data in four seasons, the proposed TNet-AIA model performs best among 10 models and has the highest prediction accuracy. On the contrary, the forecasting results of the SVM shows the lowest prediction accuracy and the worst fitting degree with actual data.

B. Compared with the four evaluation criteria across four different seasons in radar chart Fig. 11, it can be seen that the method proposed in this chapter is the best in all evaluation indicators and has the lowest prediction error. XGBoost and SVM perform poorly on the whole, especially in autumn and winter.

C. Similar to the conclusions drawn in the DKASC area, parallel TimesNet helps to enhance the forecasting accuracy. The prediction accuracy of proposed method after parallel TimesNet model is about 16.88 % higher than that of the unparallel TimesNet model.

4.5.2. Comparative analysis of stability

In order to further verify whether the proposed approach successfully enhances the Informer model's stability, we have also used the datasets of DKASC and Xuhui to conduct seasonal experimental comparisons across spring, summer, autumn and winter. To intuitively compare the stability of proposed method, we have chosen the box plot to show the experimental results.

Boxplot is a statistical tool used to display the distribution characteristics of data, which can intuitively reflect the concentration trend, dispersion degree and outliers of data. Among them, the rectangular box represents the quartile range of IQR, the black horizontal line represents the median. The boundaries of box plot indicate Q1 at the bottom and Q3 at the top. The ' whisker ' outside box is used to display the overall data range, usually extending to points within 1.5 times IQR, while points outside this interval identified as outliers (black solid dots). Through the boxplot, the error distribution of multiple models or methods can be compared, the outliers of the data can be identified, and the stability and reliability can be analyzed, providing a basis for model selection and optimization [43].

(1) DKASC area, Alice Springs, Australia

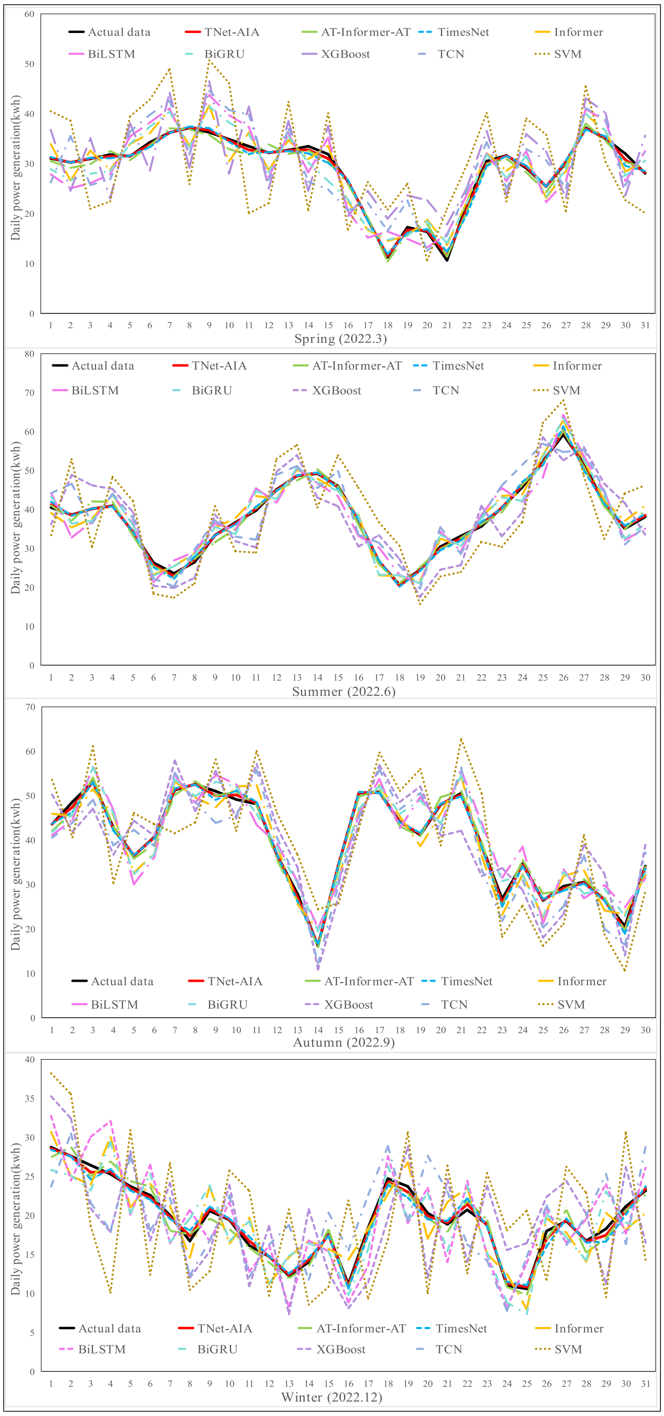


Fig. 10. Comparison between actual data and predicted results (Xuhui).

Fig. 12 illustrates the experimental results of diverse models in the DKASC region, which are analyzed in detail below:

A. In the error distribution in spring, it is apparent that the performance of different methods is quite different. Among them, the median error of TNet-AIA model has the lowest, and the overall error range is small, indicating that its prediction stability is strong. In contrast, the error distribution of SVM is the most dispersed, the median is higher, and there are multiple outliers, indicating that the prediction error is large and unstable. In addition, XGBoost and TCN also show a large error distribution, indicating that their prediction results in the quarter are not as good as other methods.
B. The error distribution trend in summer is similar to that in spring. The proposed TNet-AIA model still maintains the lowest error median and the overall error fluctuation is small. SVM is still the method with the largest error, the median is high, the distribution range is wide, and contains multiple outliers. In addition, the error distribution of XGBoost and TCN is still large, indicating that their predictive ability still has some limitations in this season. Notably, the error distribution of TimesNet, Informer, BiLSTM and BiGRU is more concentrated, indicating that they have better prediction stability in summer.
C. The error distribution in autumn still continues the trend of the previous two quarters. The proposed TNet-AIA model maintains the lowest error median and the smallest error range, indicating that its prediction stability is high. In contrast, SVM is still the most unstable method of error, showing a larger error range and more outliers. In addition, XGBoost and TCN continue to show large error distributions, while TimesNet, Informer, BiLSTM and BiGRU remain stable.
D. In winter, the median error of the proposed TNet-AIA model has still the lowest, showing the best prediction stability. SVM still has the largest error, the most widely distributed, and there are multiple outliers, which is consistent with the trend of other quarters. In addition, XGBoost and TCN still have large errors, while TimesNet, Informer, BiLSTM and BiGRU continue to maintain a relatively stable error distribution.

To better evaluate the proposed method's stability and robustness, additional experiments are conducted using multiple datasets from the Xuhui District. This regional consistency allows for a more controlled assessment of forecasting performance under varying temporal and meteorological conditions.

(2) Xuhui District, Shanghai, China

Fig. 13 presents the experimental results of diverse models in Xuhui area. The following is a detailed analysis:

A. In the prediction of spring data, the absolute error of the proposed TNet-AIA model has the lowest, the median is the smallest, and the error range is narrow, indicating that its forecasting accuracy and stability are the best. The error of AT-Infoemer-AT model is also low, and the distribution is denser, indicating that its forecasting effect is superior. The errors of TimesNet, Informer, BiLSTM and BiGRU are slightly higher, but the overall distribution is stable, indicating that their predictions in spring still have certain reliability. The error of XGBoost and TCN is large and the distribution is wide, indicating that their prediction ability in spring is weak. SVM has the highest error, the widest distribution range, and multiple outliers, indicating that its prediction error is extensive and unstable, and the forecasting performance in spring is the poorest.
B. In the prediction of summer data, the proposed TNet-AIA model still performs best, with the lowest median error and the smallest error range, indicating that the prediction effect is stable and accurate. AT-Infoemer-AT model maintains good performance, with low error, compact distribution and reliable prediction. The errors of TimesNet, Informer, BiLSTM and BiGRU are slightly higher than the first three, but they still maintain a relatively concentrated distribution, indicating that they have strong predictive stability in summer. The errors of XGBoost and TCN are relatively large and the distribution range is wide, indicating that their prediction ability decreases in summer. SVM is still the method with the highest error, the largest error range, and there are multiple outliers, indicating that its prediction performance in summer is also poor.
C. In the autumn data prediction, the proposed TNet-AIA model shows the lowest absolute error and the smallest error range, and the prediction effect is the best. AT-Infoemer-AT model maintains low error and stable distribution, indicating that they still have strong

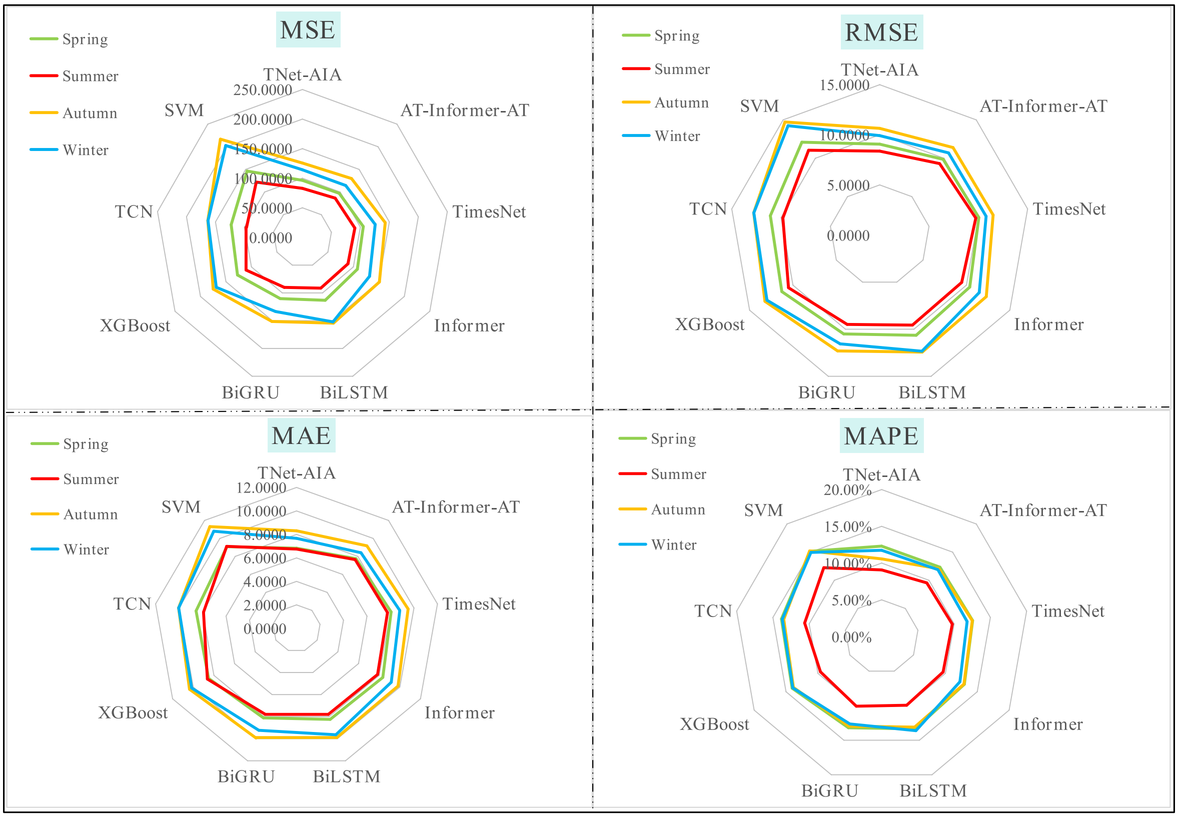


Fig. 11. Error comparison radar chart (Xuhui).

reliability in autumn data prediction. The errors of TimesNet, Informer, BiLSTM and BiGRU are slightly higher, but they still maintain a relatively concentrated distribution, and the prediction stability is better. The errors of XGBoost and TCN in autumn are still large and widely distributed, indicating that their prediction effects are weak. SVM still performs worst in autumn, with the widest error range and multiple outliers, indicating that its prediction results fluctuate greatly and have poor stability.

D. In the winter data prediction, the proposed TNet-AIA model continues to maintain the lowest absolute error, and the error range is the narrowest, showing the best forecasting precision and stability. AT-Infoemer-AT model still performs well, with low error and stable distribution, indicating that their predictive ability is still strong in winter. The errors of TimesNet, Informer, BiLSTM and BiGRU are slightly higher, but the distribution is more concentrated and the performance is moderate. XGBoost and TCN still have large errors and are widely distributed, indicating that their predictive ability is still limited in winter. SVM is still the method with the highest error, the error range is the largest, and there are multiple outliers, the worst performance, indicating that its prediction effect in winter is the least ideal.

5. Conclusion

In order to further optimize the forecasting accuracy and stability, this paper proposes a short-term PVGF method with a parallel architecture combining the strengths of TimesNet and AT-Informer-AT models. Through the experimental comparison of DKASC and Xuhui in Shanghai across the four diverse seasons, the following findings have been derived:

According to the forecasting findings of the four seasons in the two case areas, the proposed method has the best effect of fitting the actual data in the 10 models, and the forecasting precision is the highest, which is notably better than other methods. In contrast, SVM has the largest error and the weakest model fit.

Through the comparison of four evaluation metrics, the proposed method is the best in all indicators and has the lowest prediction error. XGBoost and SVM perform poorly on the whole, especially in autumn and winter.

By comparing the error distribution in different seasons, the method presented in this study has the lowest median error and smallest error range, indicating that its prediction stability is strong. XGBoost and TCN have large errors in each quarter, and SVM has the widest error distribution and the worst stability.

On the whole, the proposed TNet-AIA method demonstrates strong performance in forecasting tasks of each season, along with notable improvements in the AT-Informer-AT model's stability and the prediction accuracy of PVGF, which is increased by about 16.88 %.

Despite the improved forecasting precision achieved by the proposed PVGF method, there are still open questions worth further investigation. For instance, exploring the influence of various FE techniques in a structured way remains an important challenge. The uncertainty quantification and statistical significance analysis will be considered to more rigorously evaluate the reliability and robustness of the forecasting results. These will be the core of our future research.

CRediT authorship contribution statement

Weijie Yu: Writing – original draft, Software, Resources, Methodology, Formal analysis. **Yeming Dai:** Writing – review & editing, Supervision, Methodology, Funding acquisition. **Wenjie Wang:** Supervision, Project administration. **Tao Ren:** Visualization, Software, Data curation. **Mingming Leng:** Writing – review & editing, Supervision.

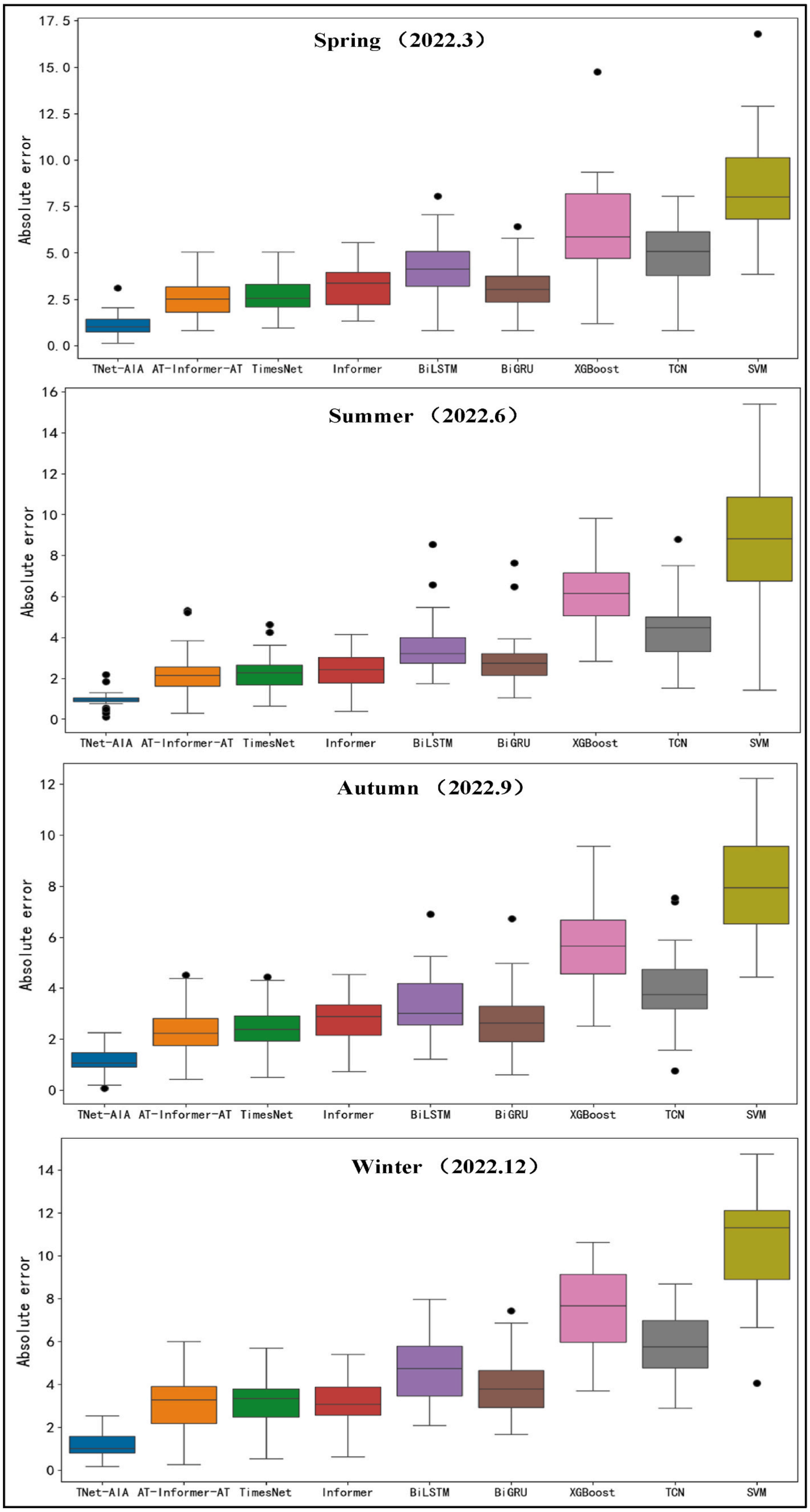


Fig. 12. Comparison of stability between models (DKASC).

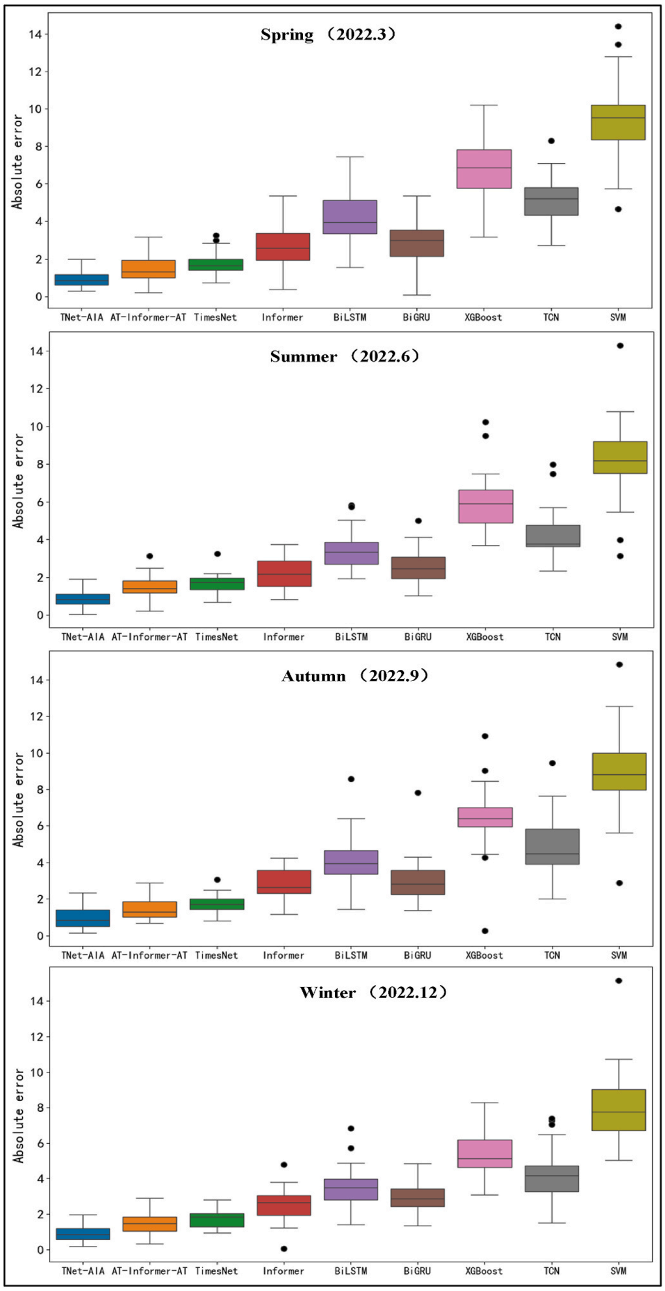


Fig. 13. Comparison of stability between models (Xuhui).

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).

References

[1] IEA, Electricity 2025, IEA, Paris, 2025. https://www.iea.org/reports/electricity-2025. Licence: CC BY 4.0.
[2] G. Piantadosi, S. Dutto, A. Galli, et al., Photovoltaic power forecasting: a transformer based framework, Energy AI 18 (2024) 10044.
[3] W.B. Qiao, L.Y. Shi, X.Z. Kong, et al., A hybrid algorithm for multi-step prediction of ultra-short-term electricity load considering an innovative preprocessing method, Renew. Energy 256 (2025) 124416.
[4] A. Meng, X. Xu, J. Chen, et al., Ultra short term photovoltaic power prediction based on reinforcement learning and combined deep learning model, Power Syst. Technol. 45 (12) (2021) 4721–4728.
[5] 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, Power Syst. Technol. 46 (10) (2022) 3940–3951.
[6] 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.
[7] 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.
[8] 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.
[9] 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.
[10] Scheinberg K. PirhooshyaranM, 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] Y. Gong, Z. Lu, Y. Qiao, et al., An overview of photovoltaic energy system output forecasting technology, Automat. Electric Sys. 40 (4) (2016) 140–151.
[16] 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.
[17] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239.
[18] 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.
[19] Z. Wang, H. Liu, S. Wu, et al., Explainable time-varying directional representations for photovoltaic power generation forecasting, J. Clean. Prod. 468 (2024) 143056.
[20] 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.
[21] 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.
[22] J. Xu, J. He, J. Gu, et al., Financial time series prediction based on XGBoost and generative adversarial networks, Int. J. Circuit. Sys. Signal Process. 16 (2022) 637–645.
[23] 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.
[24] 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.
[25] T. Peng, C. 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.
[26] Y. He, Y. Wang, Short-term wind power prediction based on EEMD-LASSO-QRNN model, Appl. Soft Comput. 105 (2021) 107288.
[27] A. Vaswani, N. Shazeer, N. Parmar, et al., Attention is all you need, Neural Inform. Processing Sys. (2017) 5998–6008.
[28] H. Zhou, S. Zhang, J. Peng, et al., Informer: beyond efficient transformer for long sequence time-series forecasting[C], Proc. AAAI Conf. Artif. Intell. 35 (12) (2021) 11106–11115.
[29] 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.
[30] M.J. Gong, Y. Zhao, J.W. Sun, et al., Load forecasting of district heating system based on informer, Energy 253 (2022) 124179.

[31] 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.
[32] 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.
[33] Z.A. Khan, T. Hussain, S.W. Baik, Dual stream network with attention mechanism for photovoltaic power forecasting, Appl. Energy 338 (2023) 120916.
[34] W.J. Yu, Y.M. Dai, T. Ren, et al., Short-time photovoltaic power forecasting based on informer model integrating attention mechanism, Appl. Soft Comput. 178 (2025) 113345.
[35] H. Wu, T. Hu, Y. Liu, et al., TimesNet: temporal 2d-variation modeling for general time series analysis, arXiv (2022), https://doi.org/10.48550/arXiv.2210.02186 preprint arXiv:221002186.
[36] H. Zhao, X.Q. Huang, Z.N. Xiao, et al., Week-ahead hourly solar irradiation forecasting method based on ICEEMDAN and TimesNet networks, Renew. Energy 220 (2024) 119706.
[37] 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.
[38] 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.
[39] 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.
[40] L. Tian, Y. Gao, H.Y. Yang, et al., Multi-scenario long-term degradation prediction of PEMFC based on generative inference informer model, Appl. Energy 377 (2025) 124398.
[41] https://www.lvsedianli.com/perHome.html.
[42] https://dkasolarcentre.com.au/download/notes-on-the-data.
[43] K. Ugur, V. Francisco, A. Murat, et al., Deep learning-based forecasting modeling of micro gas turbine performance projection: an experimental approach, Eng. Appl. Artif. Intell. 130 (2024) 107769.