ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
Received: 17 September 2025; Accepted: 5 January 2026

https://doi.org/10.17559/TV-20250917002999
Original scientific paper

TAN: A Temporal Attention Network for Photovoltaic Power Forecasting

Chao HUANG, Hui ZHANG*, Tian LUAN

Abstract: The increasing penetration of distributed photovoltaic (PV) systems poses significant challenges to power grid stability, making accurate medium and long-term PV power forecasting essential. This paper proposes a Temporal Attention Network (TAN) for multi-step PV power forecasting. TAN adopts a pure temporal self-attention architecture with a learnable temporal embedding module, enabling effective modeling of periodic and non-stationary temporal dependencies without relying on recurrent or convolutional structures. The proposed model is evaluated on real-world PV data under 24-hour and 48-hour forecasting horizons. Performance is assessed using multiple metrics, including RMSE, MAE, MAPE, R^2, and confidence intervals derived from rolling forecasting evaluation. Experimental results show that TAN achieves consistently lower forecasting errors and higher explanatory power than representative recurrent, convolutional, hybrid, and Transformer-based baselines in both forecasting settings. Moreover, TAN maintains competitive performance with moderate model complexity. These results indicate that TAN provides an effective and scalable attention-based solution for medium- and long-term PV power forecasting in renewable energy - integrated power systems.

Keywords: attention mechanism; deep learning; photovoltaic power forecasting; temporal attention network

1 INTRODUCTION

Under the "dual-carbon" strategic goal, China has introduced a series of policies, such as the Guiding Opinions on Promoting the Integration of Power Generation, Grids, Demand, and Storage and the 14th Five-Year Plan for Renewable Energy Development, to promote a high proportion of renewable energy and enhance the grid's capability to accommodate and consume new energy sources. With policy support, the coordinated development of generation, grid, load, and storage has become a key pathway toward building a new power system. As one of the fastest-growing renewable energy technologies, distributed photovoltaic (PV) systems have expanded rapidly. However, PV power generation is inherently intermittent and volatile, being strongly influenced by meteorological factors such as solar irradiance, cloud cover, and temperature. These uncertainties pose significant challenges to grid operation, especially for medium- and long-term forecasting, where insufficient accuracy may lead to source-load mismatch, voltage instability, and suboptimal energy storage scheduling.

Traditional statistical methods, such as autoregressive integrated moving average (ARIMA) [1] and exponential smoothing [2], struggle to capture the nonlinear and highly dynamic characteristics of PV power generation [3]. In recent years, deep learning models have been widely adopted for PV forecasting. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks [4], can model temporal dependencies and nonlinear dynamics but suffer from limited long-range dependency modeling and poor parallelization efficiency. Temporal convolutional networks (TCNs) [5, 6] alleviate some of these issues through dilated causal convolutions and residual connections, yet their receptive fields remain constrained by network depth and kernel size. Hybrid architectures combining TCNs with attention mechanisms [7] further improve performance but are still influenced by convolutional or recurrent inductive biases [8-11].

Motivated by the success of self-attention mechanisms, especially the Transformer architecture [12-14], in sequence modeling, this paper proposes a Temporal Attention Network (TAN) for photovoltaic power forecasting. TAN abandons recurrent and convolutional structures and employs a pure temporal self-attention mechanism to directly model dependencies among historical time steps. This design enables efficient parallel computation and flexible modeling of both short-term variations and long-term temporal patterns. Moreover, TAN is lightweight compared with standard Transformer architectures, making it suitable for deployment in distributed grid environments.

The main contributions of this paper are summarized as follows:

We propose Temporal Attention Network (TAN), a novel deep learning architecture for PV power forecasting, which replaces recurrence and convolution with a pure temporal attention mechanism.

We propose PV feature embedding (PVE). Referring to the positional encoding and word embedding in the text generation tasks, the temporal encoding and feature embedding of photovoltaic historical data are calculated to generate PV feature embeddings.

We conduct comprehensive experiments on real PV plant datasets and show that TAN outperforms existing methods (LSTM, TCN), hybrid models (TCN-Attention), Informer, and Transformer-ECA-GRU in both accuracy and stability.

2 RELATED WORKS

Photovoltaic power forecasting is an important research topic in smart grid scheduling and new energy management. With the rapid growth of photovoltaic installed capacity and the integration of source-grid-load-storage, higher demands are placed on forecasting models. These demands include accuracy, generalization ability, and stability. At present, many approaches have been adopted to build photovoltaic power forecasting models. They include traditional statistical methods, machine learning methods, and deep learning methods.

1) Statistical and physical-model-based approaches were among the earliest methods used for PV power forecasting. Linear time-series models such as ARIMA and its variants were widely applied in short-term prediction due to their simplicity, but they rely on strong stationarity assumptions and exhibit limited capability in modeling nonlinear and rapidly varying PV generation patterns [15, 16]. In parallel, several studies incorporated physical mechanisms and numerical weather prediction (NWP)

information, combining historical PV data with meteorological forecasts to estimate future power output [17, 18]. While these approaches can improve forecasting accuracy under stable weather conditions, their performance is highly dependent on the reliability of meteorological inputs and they often struggle with abrupt changes and complex nonlinear dynamics inherent in PV generation.

2) Machine learning-based methods were introduced to overcome the strong distributional assumptions and limited nonlinearity modeling capacity of traditional statistical approaches. Models such as support vector machines, ensemble learning methods (e.g., Random Forest and gradient boosting decision trees), and clustering-based frameworks were applied to PV power forecasting by leveraging historical data and engineered features [19-21]. These methods improved forecasting accuracy and robustness compared with linear models and offered better adaptability to noise and moderate variability. However, their performance still relied heavily on manual feature engineering, and they faced scalability limitations when dealing with large-scale time series and long-term temporal dependencies. In recent years, deep learning models have increasingly been adopted for PV forecasting, as they can automatically learn hierarchical feature representations and capture complex nonlinear and long-range temporal patterns directly from data.

3) Deep learning-based methods have been widely adopted for PV power forecasting due to their strong capability in modeling nonlinear temporal dynamics. Recurrent neural networks (RNNs), particularly LSTM and GRU, were extensively used to capture temporal dependencies in PV generation data and demonstrated improved robustness over traditional machine learning models [22-24]. Beyond purely recurrent architectures, convolutional neural networks (CNNs) were introduced to extract local patterns and short-term variations, and were further combined with RNNs to jointly model spatial and temporal features [25-28].

More recently, Temporal Convolutional Networks (TCNs) have attracted increasing attention in PV forecasting due to their causal structure, large receptive field enabled by dilation, and efficient parallel computation. Various TCN-based and hybrid frameworks have been proposed to enhance short-term PV prediction accuracy by integrating multiple input features or residual shrinkage mechanisms [29, 30]. In addition, combinations of RNNs and TCNs have shown promising performance in short-term forecasting tasks [31-34]. However, despite these advances, most existing deep learning approaches remain constrained by the inductive biases of recurrence or convolution, which limits their flexibility in modeling long-range temporal dependencies and complex medium- to long-term dynamics.

4) Attention mechanisms, particularly those inspired by the Transformer architecture, have recently been introduced into PV power forecasting to enhance long-range dependency modeling and parallel sequence processing [12-14]. Beyond standard Transformer-based designs, several studies proposed hybrid attention frameworks by combining attention with convolutional or recurrent backbones, such as TCN-attention models, Transformer-TCN architectures, and Transformer-RNN hybrids, in order to balance local feature extraction and long-term temporal modeling [35-37]. More recently, efficient Transformer variants, such as Informer [38], have been adopted for time-series forecasting to reduce computational overhead while handling longer sequences, and have also been explored in renewable energy forecasting scenarios.

Although these attention-based and hybrid models have shown promising results, particularly in short-term and ultra-short-term PV forecasting [39,40], they often rely on structurally complex combinations of attention, convolution, and recurrence. As a result, they inherit inductive biases from auxiliary modules and may suffer from limited scalability or reduced robustness in medium- and long-term forecasting. Developing a lightweight attention-based framework that directly models temporal dependencies without convolutional or recurrent components therefore remains an open research problem.

3 PROBLEM AND PRELIMINARIES

1) Problem: Photovoltaic (PV) power prediction problem can be transformed into a time series prediction problem. For a time series prediction problem, suppose that there is a time series $X_{t_0} = \left\{ x_{(t_0 - T_h)}, x_{(t_0 - T_h + 1)}, \cdots, x_{t_0 - 1}, x_{t_0} \right\}$, which is used to forecast next T_f -time-step states $\hat{X}_{t_0} = \left\{ \hat{x}_{t_0 + 1}, \hat{x}_{t_0 + 2}, \cdots, \hat{x}_{t_0 + T_f} \right\}$, where T_h and T_f denote the historical and predicted time window respectively, $x_t \in \mathbb{R}^{Dx}$ represents the state of the PV device at time $t \in \mathbb{R}_+$.

In the scenario of photovoltaic power prediction, the predicted state often includes only one scalar such as DC power. Thus, the problem can be simplified by using $\widehat{Y}_{t_0} = \left\{ \hat{y}_{t_0 + 1}, \hat{y}_{t_0 + 2}, \cdots, \hat{y}_{t_0 + T_f} \right\}$ representing the powers to be predicted. Now, the problem becomes to train a neural network to predict the $\widehat{Y}_{t_0}$. Suppose there is an amount of historical PV power records of a certain area or station, which record the state of PV modules of time intervals of τ denoted by $x_t = \left\{ f_t^1, f_t^2, \cdots, f_t^N \right\}$, where f_t^N denotes the *n*-th feature at time step *t*, *N* is the dimension of features, and $t \propto \tau$.

For the above problems, the mean square loss (MSE) is suitable as the loss function:

$$\min L_{\text{MSE}} = \mathbb{E}\left[\left(Y_{t_0} - \hat{Y}_{t_0} \right)^2 \right] \tag{1}$$

2) Photovoltaic Features: In (1), $\hat{Y}_t = \text{Network}\left(X_{t_0} \right)$. In this paper, for each element in X_{t_0}, the features we use include: ambient temperature, inverter temperature, module temperature, POA irradiance, relative humidity, wind direction, wind speed, month, hour, DC power. Here X_t can be expressed as:

$$x_t = \begin{Bmatrix} f_t^{\mathrm{AT}}, f_t^{\mathrm{TT}}, f_t^{\mathrm{MT}}, f_t^{\mathrm{PI}}, f_t^{\mathrm{RH}}, f_t^{\mathrm{ND}}, \\ f_t^{\mathrm{NS}}, f_t^{\mathrm{month}}, f_t^{\mathrm{pour}}, f_t^{\mathrm{DC}} \end{Bmatrix} \tag{2}$$

Most of the features can be measured directly by sensors, while the POA irradiance needs to be calculated based on some measurements. In general, three components of solar radiation can be measured directly: direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), and global horizontal irradiance (GHI). Then we can calculate POA irradiance using:

$$f^{\mathrm{PI}} = E_{\mathrm{POA}} = E_b + E_g + E_d \tag{3}$$

where E_b is the POA beam component, E_g is the POA ground-reflected component, and E_d is the POA sky-diffuse component.

The angle of incidence (AOI) between the sun's rays and the PV array can be determined as:

$$\mathrm{AOI} = \cos^{-1}\begin{bmatrix} \cos(\theta_Z)\cos(\theta_T) \\ +\sin(\theta_Z)\sin(\theta_T)\cos(\theta_A - \theta_{\mathrm{A,array}}) \end{bmatrix} \tag{4}$$

where θ_A and θ_Z are the solar azimuth and zenith angles respectively; θ_T and $\theta_{A,\mathrm{array}}$ are the tilt and azimuth angles of the array respectively. The POA beam component of irradiance is calculated by adjusting the DNI by the AOI in the following manner:

$$E_b = \mathrm{DNI} \times \cos(\mathrm{AOI}) \tag{5}$$

E_g can be calculated by:

$$E_g = \mathrm{GHI} \times albedo \times \frac{(1 - \cos(\theta_T))}{2} \tag{6}$$

where, *albedo* is the fraction of the reflected GHI. *albedo*→0 when the surface is very dark and *albedo*→1 when the surface is bright white or metallic.

E_d can be calculated by:

$$E_d = \mathrm{DHI} \times \frac{1 + \cos(\theta_T)}{2} \tag{7}$$

Thus taking (4)-(7), the POA irradiance feature (3) can be expanded as follows:

$$\begin{aligned} f^{\mathrm{PI}} &= E_b + E_g + E_d = \mathrm{DNI} \\ &\times \begin{bmatrix} \cos(\theta_Z)\cos(\theta_T) \\ +\sin(\theta_Z)\sin(\theta_T)\cos(\theta_A - \theta_{A,\mathrm{array}}) \end{bmatrix} \\ &+ \mathrm{GHI} \times albedo \times \frac{(1 - \cos(\theta_T))}{2} \\ &+ \mathrm{DHI} \times \frac{1 + \cos(\theta_T)}{2} \end{aligned} \tag{8}$$

4 METHOD

This paper focuses on photovoltaic (PV) power forecasting methods. Inspired by the success of the Transformer architecture in natural language processing (NLP) tasks, the characteristics of PV power forecasting are analyzed. Based on these characteristics, a Photovoltaic Embedding (PVE) method is designed. Furthermore, a novel network named Temporal Attention Network (TAN) is proposed.

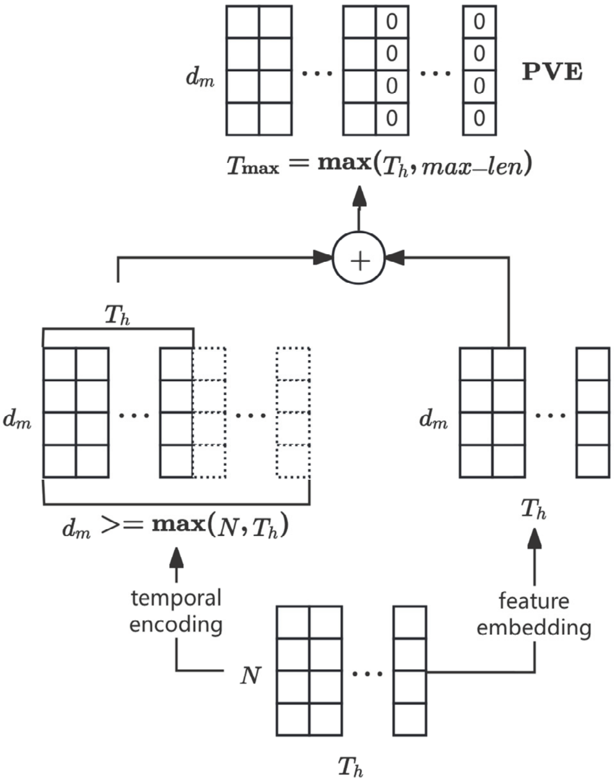


Figure 1 PV embedding

Photovoltaic Embedding. Different from text generation tasks, photovoltaic power forecasting is a time series prediction problem. In this problem, different dimensions of historical states are not represented in the same space. To address this issue, methods from word embedding and positional encoding in text generation are adopted. Feature embedding and temporal encoding are constructed jointly, referred to as PVE. The structure of PVE is illustrated in Fig. 1. As shown in text generation tasks, positional encoding models the positional information of a sequence, which can be expressed as follows [12]:

$$\begin{aligned} \mathrm{PE}_{(\mathrm{pos},2i)} &= \sin\left(pos / 10000^{2i/d_{\mathrm{model}}}\right) \\ \mathrm{PE}_{(pos,2i+1)} &= \cos\left(pos / 10000^{2i/d_{\mathrm{model}}}\right) \end{aligned} \tag{9}$$

where, *pos* is the position and i is the dimension.

In the task of medium- and long-term PV power prediction, modeling temporal characteristics is crucial. Traditional Transformer architectures usually employ fixed positional encoding to provide position-aware information at each time step. However, this method is mainly oriented toward short-term dependence modeling and shows clear limitations for periodic and nonlinear long-term relationships caused by seasonality, solar cycles, and temperature fluctuations in PV power generation. Specifically, positional encoding lacks explicit periodic representation and cannot fully capture the regularity and uncertainty of illumination and meteorological factors over time. This limitation reduces the ability of the model to perceive medium- and long-term trends. To overcome this problem, the Time2Vec method [41] is adopted to represent the temporal relationships of the input sequence.

$$\mathrm{Time2Vec}(\tau)=\begin{bmatrix}\sin\left(w_1^{\top}\tau+b_1\right),\cdots,\\ \sin\left(w_k^{\top}\tau+b_k\right),w_0^{\top}\tau+b_0\end{bmatrix} \tag{10}$$

Feature embedding maps the input features to an embedding space through a linear layer:

$$\mathrm{FE}=\sigma\left(W_{\theta}^{\mathrm{FE}}X_{t_0}+b_{\theta}^{\mathrm{FE}}\right) \tag{11}$$

where, σ is the activation function.

The input features of TAN are obtained by adding the temporal encoding (TE) and the feature embedding (FE):

$$X^{\mathrm{PVE}}=\mathrm{ZeroPad}\left(\mathrm{FE}+\mathrm{TE}\right) \tag{12}$$

TAN. For time series problems, models from the RNN family and TCN are most commonly used. However, RNNs are limited by the difficulty of parallel computing. TCNs usually employ small windows (e.g., 3 or 5), which made it difficult to ensure stability in long-term sequence generation. To address these issues, a Temporal Attention Network (TAN) is proposed, which integrates the self-attention mechanism that has achieved great success in text generation tasks in recent years. TAN is designed to forecast photovoltaic power with improved efficiency and stability.

The self-attention mechanism was introduced into sequence prediction and has achieved widespread success in recent years. Compared with RNNs and CNNs, this mechanism is convenient for parallel computation and easy to stack. These advantages have led to significant achievements in many sequence prediction tasks. The attention mechanism takes an input matrix $X\in\mathbb{R}^{\boldsymbol{T}\times\boldsymbol{D}}$, where T denotes the length of the input sequence. In the PV power prediction problem, T corresponds to the length of the historical time series of the input states. D represents the dimension of each state. The matrices Q, K and V are obtained through different linear transformations of X. The attention of the input is then calculated as follows [12]:

$$\mathrm{Attention}\left(Q,K,V\right)=\mathrm{softmax}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{13}$$

where d_k indicates dimension of K.

Based on the attention mechanism described above, a Temporal Attention Network (TAN) is proposed to analyze and forecast time-series photovoltaic power. The overall architecture of TAN is illustrated in Fig. 2.

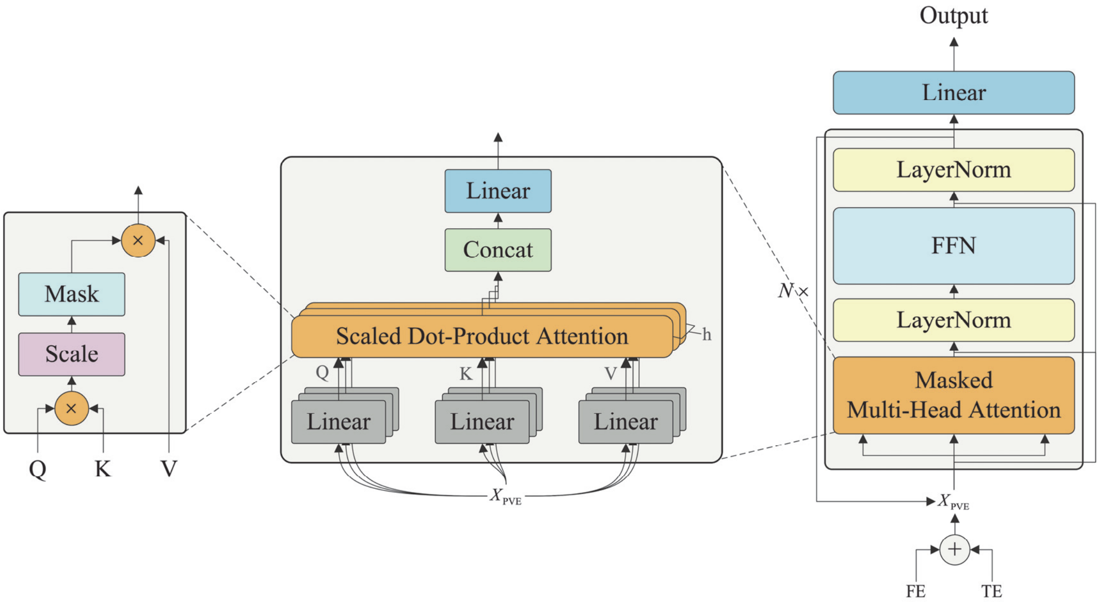


Figure 2 TAN structure

As described in the previous section, and similar to positional encoding in text generation tasks, the proposed model does not contain recurrent or convolutional modules. Therefore, PVE is employed to represent the temporal relationships of the input sequence, as defined in (12).

Within each residual block, the input sequence X^{PVE} obtained after photovoltaic embedding is multiplied by W_Q,W_K,W_V to obtain Q, K, V, respectively. The attention formulation in (13) is then modified to incorporate the mask mechanism (MMHA) for the time series prediction task:

$$\mathrm{Attention}\left(Q,K,V\right)=M_{\mathrm{temporal}}\left(\frac{QK^{T}}{\sqrt{d_k}}\right)V \tag{14}$$

where $M_{\text{temporal}} \in \mathbb{R}^{d_k \times d_k}$ is an upper triangular matrix representing the temporal causal mask.

The structure of TAN is illustrated in Fig. 2 and is composed of multiple residual blocks. Each block mainly contains multi-head self-attention, a feed-forward network (FFN), and layer normalization. In the attention module, the multi-head self-attention mechanism is applied to extract the PV state time series features X^{PVE}. Compared with the traditional attention mechanism, multi-head attention divides the representation of the input sequence into multiple subspaces (heads) and calculates attention weights independently in each subspace. The results of these subspaces are then concatenated, enabling the model to capture different contextual information across subspaces. The concatenated outputs are processed by layer normalization and passed into the FFN. Finally, the output of each block is obtained by combining the block input (residual connection) with the transformed representation.

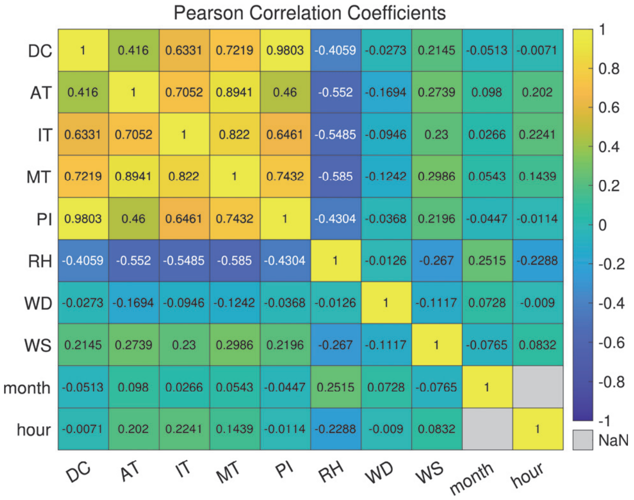
$$\text{Output}_{\text{ResBlock}} = \text{LN}\left(\text{FFN}\left(\text{MMHA}\left(X^{\text{PVE}}\right)\right) + X^{\text{PVE}}\right) \tag{15}$$

After multiple residual blocks, two linear layers are added to generate the predicted PV power, denoted as $\hat{Y}_{t_0} = \left\{\hat{y}_{t_0+1}, \hat{y}_{t_0+2}, \cdots, \hat{y}_{t_0+T}\right\}$ as defined in (16).

$$\hat{Y} = ReLU\left(\text{ Linears}\left(\text{Output}_{ResBlock}\right)\right)\left[\,:\,,\,:T_f\,,-1\right] \tag{16}$$

5 EXPERIMENTS

This section presents the experimental study. The first subsection introduces the selected dataset and the preprocessing procedures. The second subsection describes the hyper parameter settings of the model used in the experiments. The third subsection reports the experimental results and provides corresponding analysis.



	DC	AT	IT	MT	PI	RH	WD	WS	month	hour
DC	1	0.416	0.6331	0.7219	0.9803	-0.4059	-0.0273	0.2145	-0.0513	-0.0071
AT	0.416	1	0.7052	0.8941	0.46	-0.552	-0.1694	0.2739	0.098	0.202
IT	0.6331	0.7052	1	0.822	0.6461	-0.5485	-0.0946	0.23	0.0266	0.2241
MT	0.7219	0.8941	0.822	1	0.7432	-0.585	-0.1242	0.2986	0.0543	0.1439
PI	0.9803	0.46	0.6461	0.7432	1	-0.4304	-0.0368	0.2196	-0.0447	-0.0114
RH	-0.4059	-0.552	-0.5485	-0.585	-0.4304	1	-0.0126	-0.267	0.2515	-0.2288
WD	-0.0273	-0.1694	-0.0946	-0.1242	-0.0368	-0.0126	1	-0.1117	0.0728	-0.009
WS	0.2145	0.2739	0.23	0.2986	0.2196	-0.267	-0.1117	1	-0.0765	0.0832
month	-0.0513	0.098	0.0266	0.0543	-0.0447	0.2515	0.0728	-0.0765	1	
hour	-0.0071	0.202	0.2241	0.1439	-0.0114	-0.2288	-0.009	0.0832		1

Figure 3 Pearson correlation coefficient between features

Dataset and Preprocessing. To evaluate the effectiveness of the proposed TAN, experiments are conducted using the public PVDAQ dataset. PVDAQ is a large-scale time-series database containing system metadata and performance measurements from experimental, residential, and commercial PV sites in the United States. In this study, we use hourly-resolution data covering a continuous three-year period from 2012 to 2014, which includes complete seasonal cycles. The recorded variables include DC power output, ambient temperature, inverter temperature, module temperature, plane-of-array (POA) irradiance, relative humidity, wind direction, and wind speed. Fig. 3 illustrates the Pearson correlation coefficients among these features.

Prior to model training, raw measurements were organized into multivariate time series and arranged in a uniform hourly grid. Negative DC power values due to measurement noise are clipped to zero to ensure physical consistency. To reduce scale differences between heterogeneous sensor channels, feature normalization is applied by scaling the maximum absolute magnitude of each variable across the training range. Supervised learning samples are constructed using a sliding window strategy: historical observations over a window of length L are used to predict subsequent DC power sequences of H steps. To evaluate the mid to long term forecasting performance, two forecasting Settings are considered, namely $L = H = 24$ (1 day ahead) and $L = H = 48$ (2 days ahead). The generated samples are divided into training set, validation set and test set in chronological order, with a ratio of 6:2:2 to ensure strict temporal order and avoid information leakage in future time periods.

Implementations. In the PV power prediction task, the proposed TAN is compared with several representative models, which show strong performance in time series prediction. These baselines include classical recurrent and convolutional models (LSTM, TCN), hybrid convolutional Attention models (TCN-Attention), and two recent attention-based architectures, Namely Informer [38] and a hybrid Transformer-GRU model augmented with Efficient Channel Attention (ECA) (Transformer-ECA-GRU), which adds both architectures to the revised experiments to strengthen the baseline comparison. For fair comparison, the number of training epochs was fixed to 20 for all models.

Table 1 Model parameters

Model	dim_model	dim_channel	num_block	num_head
LSTM	256	\	3	\
TCN	\	64	3	\
TCN-Attention	512	64	3	1
Informer	128	\	4	8
Transformer-ECA-GRU	128 (64)	\	2	8
TAN24	24	\	6	8
TAN48	48	\	6	8

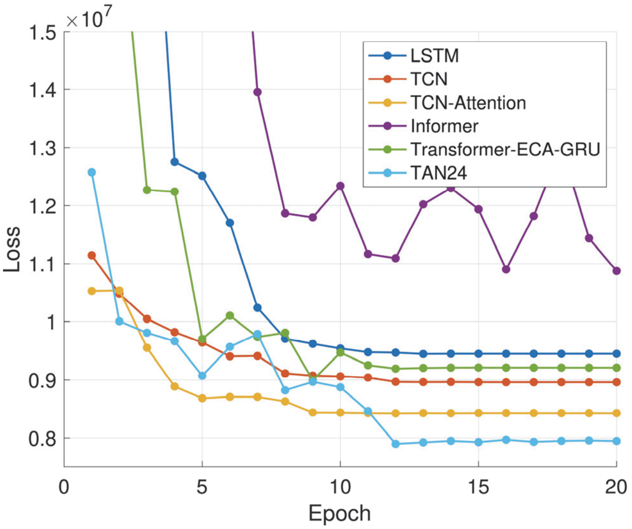
For LSTM, three LSTM layers are used, and each layer has a hidden size of 256. For TCN, three causal dilated convolutional layers are adopted, each with a kernel size of 3. The dilation factors are set to 1, 2, and 4, respectively, and the number of channels in the hidden layers is 64. For TCN-Attention, the structure is similar to TCN with three causal dilated convolutional layers. In addition, the model dimension of the attention module is set to 512. After tuning, Informer is set to 4 attention blocks and model dimension is set to 128. Transformer-ECA-GRU combines self-attention, efficient channel attention (ECA), and gated recurrent units (GRU), with a model dimension of 128 (64 for GRU), 2 Transformer blocks, and 1 GRU cell. For the proposed TAN, six stacked multi-head masked Self-attention (MMHA) residual blocks are employed. The model dimension was set to 24 for the 24-step prediction task and 48 for the 48-step prediction task, with 8 attention heads in both cases. Here model dimension "dim_model" denotes the latent dimensionality used in attention computation and feature representation. All model configurations are summarized in Tab. 1.

Results Fig. 4 shows the training loss curves of the six compared models under two multi-step forecasting settings. Fig. 4i presents the results for the 24-step forecasting task, while Fig. 4ii illustrates the results for the 48-step forecasting task.

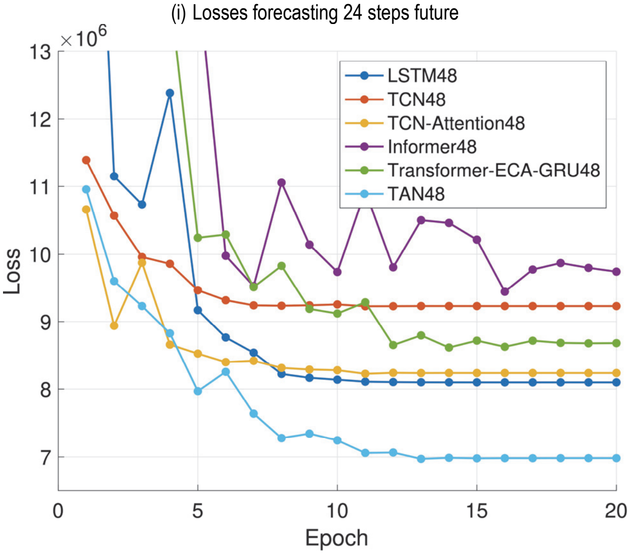
As shown in Fig. 4i, clear differences in convergence behavior can be observed among the models in the 24-step forecasting task. The LSTM model converges slowly and stabilizes at a relatively high loss level. The TCN and TCN-Attention models achieve faster convergence and lower final losses than LSTM, indicating improved temporal modeling capability. The Transformer-ECA-GRU hybrid model exhibits moderate convergence performance, outperforming LSTM but remaining inferior to TCN-Attention. In contrast, the Informer model shows pronounced instability and maintains the highest loss values throughout training, with significant oscillations across epochs. The proposed TAN achieves the lowest final training loss and demonstrates stable convergence after approximately 10 epochs.

For the more challenging 48-step forecasting task shown in Fig. 4ii, the performance gap among models becomes more evident. The Informer model continues to exhibit large fluctuations and the highest loss values, indicating difficulty in optimization under longer forecasting horizons. LSTM shows noticeable variance in early epochs and converges to a higher loss compared with convolution-based and attention-based models. The TCN and TCN-Attention models maintain relatively stable convergence, with TCN-Attention consistently outperforming TCN. The Transformer-ECA-GRU model improves stability compared to Informer but still converges to a higher loss than TAN. Notably, the proposed TAN achieves both the fastest convergence and the lowest final loss among all models, with minimal variance after convergence.

Tab. 2 and Tab. 3 report the quantitative comparison results of all six models under 24-step and 48-step forecasting settings, respectively. Overall, the proposed TAN achieves the lowest RMSE and MAE, as well as the highest R^2, in both forecasting horizons, indicating superior accuracy and explanatory capability. In the 24-step task, TAN24 reduces RMSE compared with recurrent, convolutional, hybrid, and Transformer-based baselines, while also exhibiting narrower prediction uncertainty as reflected by the CI85, CI90, and CI95 intervals. For the more challenging 48-step forecasting task, performance differences become more pronounced: although LSTM and TCN-based models show competitive short-range behavior, their errors increase noticeably with longer horizons, whereas TAN48 maintains stable accuracy and achieves the best overall performance across most metrics. The Informer model consistently yields substantially larger errors and wider confidence intervals in both settings, suggesting limited suitability for this moderate-length PV forecasting scenario. The Transformer-ECA-GRU hybrid model improves stability relative to Informer but remains inferior to TAN, particularly in long-horizon prediction. These results demonstrate that TAN provides more accurate and robust medium- and long-term PV power forecasts, which is consistent with the observed convergence behavior in Fig. 4.



(i) Losses forecasting 24 steps future



(ii) Losses forecasting 48 steps future.

Figure 4 Training result

To further illustrate the qualitative forecasting performance, five representative samples (corresponding to five days) are randomly selected from the test set, and the prediction results of six models are depicted in Fig. 5i-vi. The ground truth PV power is shown in blue, while the predicted outputs are indicated by red curves. As shown in Fig. 5i, the LSTM model exhibits obvious over-smoothing. The systematic underestimation near the peak and the lagged response to the rapid rise and fall phases within the day reflect its limited ability to characterize high-frequency changes in medium and long term modeling. Figs. 5ii-iii show that TCN and TCN-Attention are more sensitive to local fluctuations than LSTM and can partially capture the power change trend, but there is still a large amplitude deviation in the stage of rapid power change or peak, especially in the high fluctuation interval. Fig. 5v shows that Informer produces obvious flat top or oscillation prediction results in multiple time periods and fails to accurately align the true peak position, indicating that it has the problems of training instability and insufficient generalization ability under the condition of current data scale and non-stationary PV sequence. In contrast, Transformer-ECA-GRU (Fig. 5vi) performs more robustly in overall trend modeling, but there is still a certain phase offset and local overfitting in the abrupt interval.

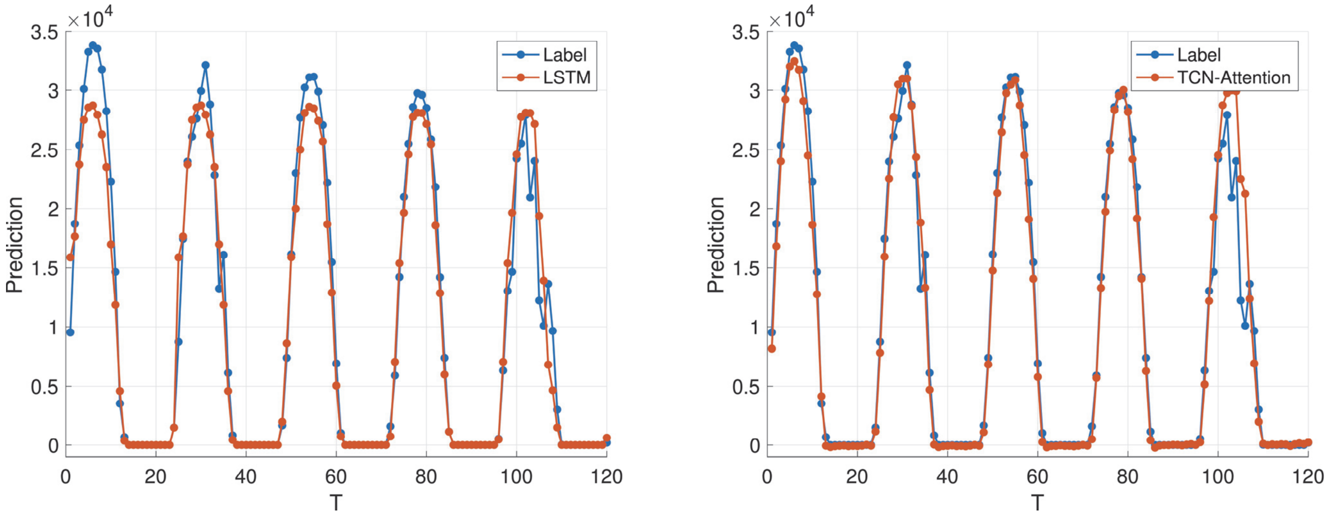
As shown in Fig. 5iv, the proposed TAN maintains the closest agreement with the true power curve in all samples, which can not only accurately characterize the intra-day periodic patterns, but also maintain stable prediction performance under cloud cover changes or severe power fluctuations. This advantage mainly benefits from the attention mechanism's ability to model the dynamic dependence of the full history time step, and Time2Vec's effective representation of periodic and non-stationary time features. On the whole, the visualization results of Fig. 5 are highly consistent with the quantitative assessment results mentioned above, which further verifies the accuracy and robustness of TAN in medium and long-term PV power prediction tasks, and shows its application potential in actual complex operation scenarios.

Table 2 Comparisons of models - 24 steps

Model	RMSE	MAE	MAPE	R^2	CI85	CI90	CI95
TAN24	**2828.57**	**1177.25**	**622061888.00%**	**0.9327**	**4705.54**	**7130.54**	**12879.49**
LSTM [4]	3009.77	1417.91	740607872.00%	0.9238	5896.75	8568.86	13970.41
TCN [5]	2961.79	1285.50	3697115648.00%	0.9262	4822.70	7772.74	14379.15
TCN-Attention [7]	3075.02	1383.04	5614354944.00%	0.9204	4934.96	7691.26	14640.70
Informer [38]	5675.25	1441.09	1190606336.00%	0.9103	5648.92	8716.77	15257.96
Transformer-ECA-GRU	3019.57	1247.72	1416622720.00%	0.9233	4807.73	7695.03	14420.29

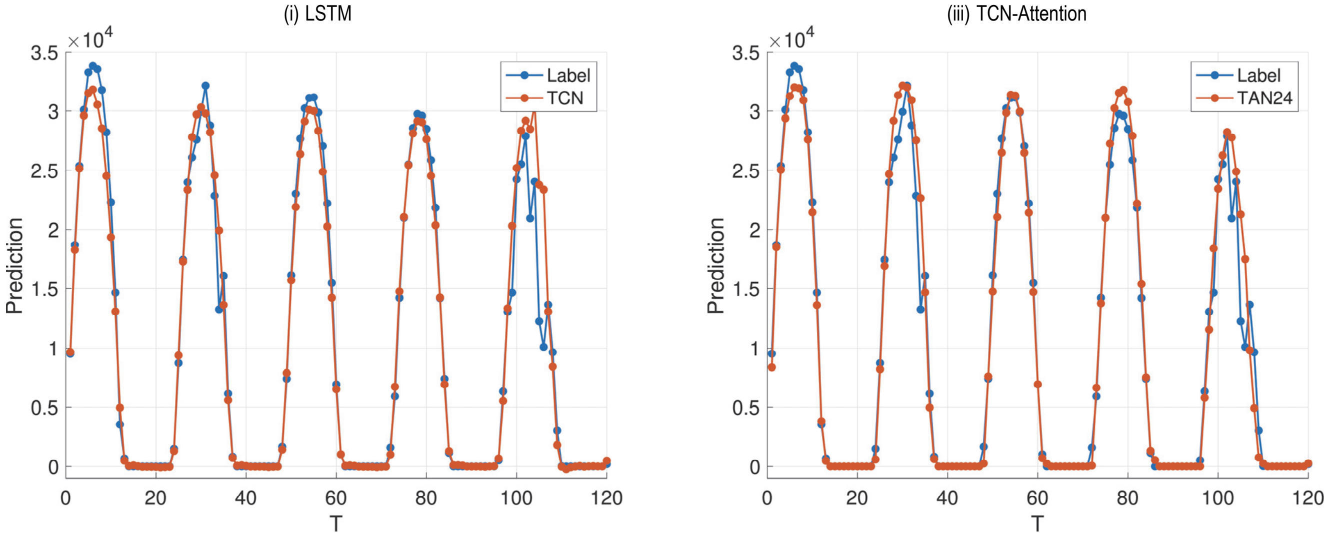
Table 3 Comparisons of models - 48 steps

Model	RMSE	MAE	MAPE	R^2	CI85	CI90	CI95
TAN48	**2812.82**	**1218.16**	**734060096.00%**	**0.9334**	5043.71	**7368.91**	**12329.73**
LSTM [4]	3230.92	1393.59	1272167936.00%	0.9122	6291.02	9040.04	15208.89
TCN [5]	3101.60	1364.60	4558746624.00%	0.9191	5391.36	8586.82	14810.77
TCN-Attention [7]	3054.16	1352.88	5980147712.00%	0.9215	**4823.50**	7897.32	14542.23
Informer [38]	3909.91	1896.82	2466552832.00%	0.8714	8619.27	12359.06	19152.27
Transformer-ECA-GRU	3205.42	1406.73	2527375360.00%	0.9135	6220.76	8867.93	14859.45



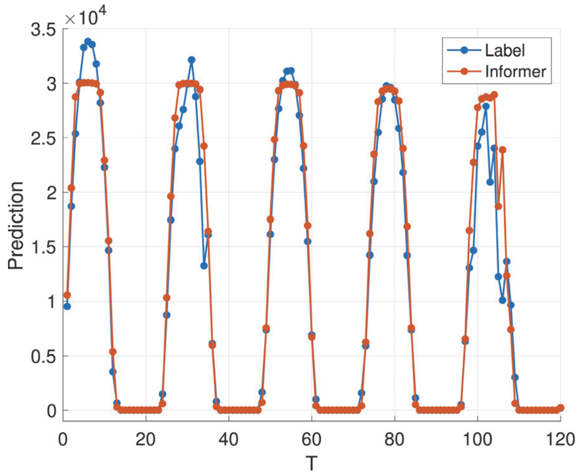
(i) LSTM

(iii) TCN-Attention

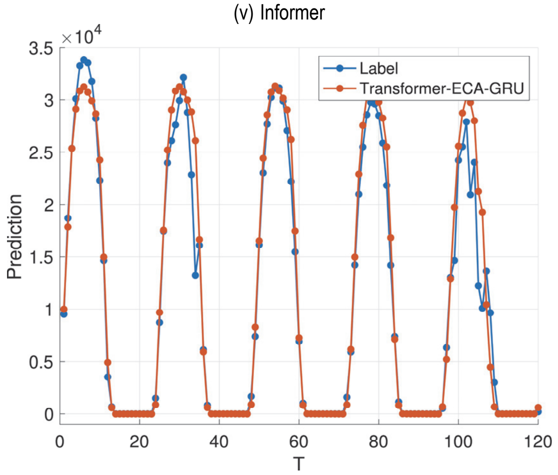


(ii) TCN

(iv) TAN24



(v) Informer



(vi) Transformer-ECA-GRU

Figure 5 5 days 5*24 h forecasting - blue lines are labels, red lines are forecasts

6 CONCLUSIONS

This paper proposes a Temporal Attention Network (TAN) for short- to medium-term photovoltaic (PV) power forecasting. By leveraging a fully attention-based architecture combined with learnable temporal encoding, TAN is able to explicitly model long-range temporal dependencies and dynamic temporal importance across historical observations. Compared with conventional sequence models such as LSTM, TCN, and hybrid TCN-attention or Transformer-RNN structures, TAN demonstrates stronger representation capability for capturing periodicity, non-stationarity, and abrupt variations inherent in PV generation data.

Comprehensive experiments conducted on real-world PV datasets validate the effectiveness of the proposed approach. Across multiple forecasting horizons (24 h and 48 h) and evaluation metrics, including RMSE, MAE, R^2, and prediction interval coverage, TAN consistently achieves superior performance over all baseline models. Both quantitative results and qualitative visualizations indicate that TAN exhibits improved robustness under highly fluctuating generation conditions and maintains stable performance in longer forecasting horizons, highlighting the advantages of attention-based global dependency modeling in PV power prediction.

Future work will focus on extending TAN to larger and more diverse datasets, incorporating data from multiple geographical locations, climatic zones, and different types of PV systems. Additional contextual information, such as geographical and meteorological attributes, will be integrated to further enhance model generalization. Moreover, combining TAN with physics-informed modeling and hybrid learning frameworks will be explored to improve interpretability and reliability, facilitating practical deployment in real-world power system operation and energy management scenarios.

Acknowledgement

This paper is supported by the project "Key Technology Research and Product Development of digital Enhancement of New active Distribution Network with Source Network load and storage Coordination" of Beijing Fibrlink Communications Co., LTD. (Project number: 546826240038.

7 REFERENCES

[1] Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. *Future Internet*, *15*(8), 255. https://doi.org/10.3390/fi15080255

[2] Iheanetu, K. J. (2022). Solar photovoltaic power forecasting: A review. *Sustainability*, *14*(24), 17005. https://doi.org/10.3390/su142417005

[3] Wang, S., Li, C., & Lim, A. (2019). Why are the ARIMA and SARIMA not sufficient. *arXiv preprint arXiv:1904.07632*. https://doi.org/10.48550/arXiv.1904.07632

[4] Dakheel, F. & Çevik, M. (2025). Electric Load Forecasting and Management in Smart Grids Using Optimized Long Short-Term Memory Network: A Real-World Evaluation. *International Journal of Energy Production and Management*, *10*(3), 371-383. https://doi.org/10.56578/ijepm100302

[5] Lea, C., Vidal, R., Reiter, A., & Hager, G. D. (2016). Temporal convolutional networks: A unified approach to action segmentation. In *European Conference on Computer Vision,* 47-54. https://doi.org/10.1007/978-3-319-49409-8_7

[6] Wang, T. C., Teng, Q. J., & Jin, G. H. (2024). A Remaining Useful Life Prediction Method for Rolling Bearings Based on Broad Learning System - Multi-Scale Temporal Convolutional Network. *Precision Mechanics & Digital Fabrication*, *1*(3), 145-157. https://doi.org/10.56578/pmdf010303

[7] Hao, H., Wang, Y., Xue, S., Xia, Y., Zhao, J., & Shen, F. (2020). Temporal convolutional attention-based network for sequence modeling. *arXiv preprint arXiv:2002.12530.* https://doi.org/10.48550/arXiv.2002.12530

[8] Kharitonov, E. & Chaabouni, R. (2020). What they do when in doubt: a study of inductive biases in seq2seq learners. *arXiv preprint arXiv:2006.14953*. https://doi.org/10.48550/arXiv.2006.14953

[9] Gunasekar, S., Lee, J. D., Soudry, D., & Srebro, N. (2018). Implicit bias of gradient descent on linear convolutional networks. *Advances in Neural Information Processing Systems*.

[10] Wang, Z., Wu, L. (2023). Theoretical analysis of the inductive biases in deep convolutional networks. *Advances in Neural Information Processing Systems*, *36*, 74289-74338. https://doi.org/10.52202/075280-3249

[11] Tharani, P. P. & Baranidharan, b. (2024). A Hybrid ViT-CNN Model Premeditated for Rice Leaf Disease Identification. *International Journal of Computational Methods and Experimental Measurements*, *12*(1), 35-43. https://doi.org/10.18280/ijcmem.120104

[12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, 6000-6010.

[13] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies,* Minnesota, 4171-4186. https://doi.org/10.18653/v1/N19-1423

[14] Peng, S. Q., Xi, G. Q., Wei, Y. S., & Yu, L. (2024). Enhanced Detection of Soybean Leaf Diseases Using an Improved Yolov5 Model. *International Journal of Knowledge and Innovation Studies*, *2*(1), 45-56. https://doi.org/10.56578/ijkis020105

[15] Fara, L., Diaconu, A., Craciunescu, D., & Fara, S. (2021). Forecasting of energy production for photovoltaic systems based on ARIMA and ANN advanced models. *International Journal of Photoenergy*, *2021*(1), 6777488. https://doi.org/10.1155/2021/6777488

[16] Li, Y., Su, Y., & Shu, L. (2014). An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, *66*, 78-89. https://doi.org/10.1016/j.renene.2013.11.067

[17] Monteiro, C., Santos, T., Fernandez-Jimenez, L. A., Ramirez-Rosado, I. J., & Terreros-Olarte, M. S. (2013). Short-term power forecasting model for photovoltaic plants based on historical similarity. *Energies*, *6*(5), 2624-2643. https://doi.org/10.3390/en6052624

[18] Fernandez-Jimenez, L. A., Muñoz-Jimenez, A., Falces, A., Mendoza-Villena, M., Garcia-Garrido, E., Lara-Santillan, P. M., & Zorzano-Santamaria, P. J. (2012). Short-term power forecasting system for photovoltaic plants. *Renewable Energy*, *44*, 311-317. https://doi.org/10.1016/j.renene.2012.01.108

[19] Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Transactions on Industry Applications*, *48*(3), 1064-1069. https://doi.org/10.1109/TIA.2012.2190816

[20] Liu, D. & Sun, K. (2019). Random forest solar power forecast based on classification optimization. *Energy*, *187*, 115940. https://doi.org/10.1016/j.energy.2019.115940

[21] Wang, J., Li, P., Ran, R., Che, Y., & Zhou, Y. (2018). A short-term photovoltaic power prediction model based on the gradient boost decision tree. *Applied Sciences*, *8*(5), 689. https://doi.org/10.3390/app8050689

[22] Abdel-Nasser, M. & Mahmoud, K. (2019). Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Computing and Applications*, *31*(7), 2727-2740. https://doi.org/10.1007/s00521-017-3225-z

[23] Konstantinou, M., Peratikou, S., & Charalambides, A. G. (2021). Solar photovoltaic forecasting of power output using LSTM networks. *Atmosphere*, *12*(1), 124. https://doi.org/10.3390/atmos12010124

[24] Akhter, M. N., Mekhilef, S., Mokhlis, H., Almohaimeed, Z. M., Muhammad, M. A., Khairuddin, A. S. M., & Hussain, M. M. (2022). An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants. *Energies*, *15*(6), 2243. https://doi.org/10.3390/en15062243

[25] Venugopal, V., Sun, Y., & Brandt, A. R. (2019). Short-term solar PV forecasting using computer vision: The search for optimal CNN architectures for incorporating sky images and PV generation history. *Journal of Renewable and Sustainable Energy*, *11*(6), 066102. https://doi.org/10.1063/1.5122796

[26] Wang, K., Qi, X., & Liu, H. (2019). Photovoltaic power forecasting based LSTM-Convolutional Network. *Energy*, *189*, 116225. https://doi.org/10.1016/j.energy.2019.116225

[27] Agga, A., Abbou, A., Labbadi, M., & El Houm, Y. (2021). Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. *Renewable Energy*, *177*, 101-112. https://doi.org/10.1016/j.renene.2021.05.095

[28] Suresh, V., Janik, P., Rezmer, J., & Leonowicz, Z. (2020). Forecasting solar PV output using convolutional neural networks with a sliding window algorithm. *Energies*, *13*(3), 723. https://doi.org/10.3390/en13030723

[29] Li, Y., Song, L., Zhang, S., Kraus, L., Adcox, T., Willardson, R., & Lu, N. (2023). A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. *IEEE Transactions on Smart Grid*, *14*(5), 4073-4085. https://doi.org/10.1109/TSG.2023.3236992

[30] Wang, M., Rao, C., Xiao, X., Hu, Z., & Goh, M. (2024). Efficient shrinkage temporal convolutional network model for photovoltaic power prediction. *Energy*, *297*, 131295. https://doi.org/10.1016/j.energy.2024.131295

[31] Limouni, T., Yaagoubi, R., Bouziane, K., Guissi, K., & Baali, E. H. (2023). Accurate one step and multistep forecasting of very short-term PV power using LSTM-TCN model. *Renewable Energy*, *205*, 1010-1024. https://doi.org/10.1016/j.renene.2023.01.118

[32] Choumal, A., Rizwan, M., & Jha, S. (2024). LSTM and TCN based Hybrid Networks for Probabilistic Photovoltaic Power Prediction. *2024 International Conference on Modeling, Simulation & Intelligent Computing (MoSICom)*, 254-259. https://doi.org/10.1109/MoSICom63082.2024.10881676

[33] Zhang, W., Zhong, A., Duan, K., & Shao, L. (2024). Ultra-Short-term load forecasting using TCN-LSTM with photovoltaic penetration rate and temporal feature extraction. In *2024 IEEE 8th Conference on Energy Internet and Energy System Integration (EI2)*, 3228-3236. https://doi.org/10.1109/EI264398.2024.10991806

[34] Zhang, X. (2023). Research on photovoltaic power prediction method based on TCN-BiLSTM neural network. *2023 IEEE 6th International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, 734-740. https://doi.org/10.1109/AUTEEE60196.2023.10407564

[35] Zhu, K. & Li, W. (2024). Short-term PV power prediction method based on weather classification and TCN-attention. *2024 8th International Conference on Smart Grid and Smart Cities (ICSGSC)*, 153-158. https://doi.org/10.1109/ICSGSC62639.2024.10813753

[36] Wu, L., Yu, J., Dai, Y., Gao, T., & Zhang, J. (2024). Photovoltaic power generation forecasting based on TCN-transformer model. *2024 5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*, 620-626. https://doi.org/10.1109/AIEA62095.2024.10692906

[37] Xiang, X., Li, X., Zhang, Y., & Hu, J. (2024). A short-term forecasting method for photovoltaic power generation based on the TCN-ECANet-GRU hybrid model. *Scientific Reports*, *14*(1), 6744. https://doi.org/10.1038/s41598-024-56751-6

[38] Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021, May). Informer: Beyond efficient transformer for long sequence time-series forecasting. *Proceedings of the AAAI conference on artificial intelligence*, *35*(12), 11106-11115. https://doi.org/10.1609/aaai.v35i12.17325

[39] Lin, H., Gao, L., Cui, M., Liu, H., Li, C., & Yu, M. (2025). Short-term distributed photovoltaic power prediction based on temporal self-attention mechanism and advanced signal decomposition techniques with feature fusion. *Energy*, *315*, 134395. https://doi.org/10.1016/j.energy.2025.134395

[40] Zhou, Z., Dai, Y., & Leng, M. (2025). A photovoltaic power forecasting framework based on Attention mechanism and

parallel prediction architecture. *Applied Energy*, *391*, 125869. https://doi.org/10.1016/j.apenergy.2025.125869

[41] Kazemi, S. M., Goel, R., Eghbali, S., Ramanan, J., Sahota, J., Thakur, S., & Brubaker, M. (2019). Time2vec: Learning a vector representation of time. *arXiv preprint arXiv:1907.05321*. https://doi.org/10.48550/arXiv.1907.05321

Contact information:

Chao HUANG
Beijing Fiberlink Communications Co., LTD.,
Beijing 100010, China
E-mail: huangchao@sgitg.sgcc.com.cn

Hui ZHANG
(Corresponding author)
Beijing Fiberlink Communications Co., LTD.,
Beijing 100010, China
E-mail: zhanghui_zdfh@163.com

Tian LUAN
Beijing Fiberlink Communications Co., LTD.,
Beijing 100010, China
E-mail: luantian@sgitg.sgcc.com.cn