

RESEARCH ARTICLE

A Two-Stage Method for Ultra-Short-Term PV Power Forecasting Based on Data-Driven

HANGXIA ZHOU¹, JUN WANG¹, FULIAN OUYANG¹, CHEN CUI², AND XIANBIN LI¹¹Laboratory of Big Data and Artificial Intelligence, College of Information and Engineering, China Jiliang University, Hangzhou 310018, China²Key Laboratory of Public Security Information Application Based on Big Data Architecture, Zhejiang Police College, Hangzhou 310053, China

Corresponding author: Jun Wang (s20030812009@cjlu.edu.cn)

This work was supported in part by the Zhejiang Province Public Welfare Technology Application Research Project under Grant LGG22E070003, and in part by the Ministry of Public Security Open Subjects of the Key Laboratory of Public Security Information Application Based on Big Data Architecture under Grant 2021DSJSYS004.

ABSTRACT To promote the real-time dispatching of a power grid and balanced decision-making of power producers, accuracy and real-time forecasting are two main problems that need to be solved in ultra-short-term photovoltaic (PV) forecasting. Focusing on the problems of slow model training speed and low forecasting accuracy due to the redundancy of training data samples and insufficient long periodic capture of data in complex weather, this paper proposes a two-stage method for ultra-short-term PV power forecasting based on data-driven. In the meteorological analysis stage, the generation power samples similar to the forecast day were extracted by inputting daily meteorological features and using maximal information coefficient (MIC) weighted grey correlation degree to form the corresponding forecast data set. In the power forecasting stage, the temporal convolutional network (TCN) extracts local features to maintain the sequence of extracted features. Then the bidirectional gating unit (BiGRU) combined with the Skip connection strategy was used to fully learn the long and the short time sequence of photovoltaic sequences, and the attention mechanism was used to pay adaptive attention to the more important historical states. This study experimented on the measured data from a photovoltaic power station in Southeastern China. The experimental results show that this method is effective in photovoltaic power short-term forecasts. In addition, compared with the latest model, this method has smaller forecasting errors and higher robustness in the time scales of 15 minutes, 30 minutes, 45 minutes, and 60 minutes. Specifically, indicator R2 increased by an average of 5.4%, while indicators RMSE and MAE decreased by 8.6% and 7.3%, respectively.


INDEX TERMS Bidirectional gating unit, temporal convolutional network, maximal information coefficient, skip connection, photovoltaic power forecasting.

I. INTRODUCTION

Due to the influence of factors such as solar radiation and temperature, PV power generation exhibits temporal variability and unpredictability, which can pose a threat to the stability and security of electric systems [1]. PV power forecasting can be broadly classified into four categories based on the time scale involved [2]: medium-term (a week to a few months) and long-term (1 to 10 years) power generation forecasting is mainly used for long-term planning of power grids and power station operation management; short-term forecasting

(1h to 1 week) can promote economic load distribution and power system operation. Ultra-short-term forecasting (within 1h) is of great significance in power marketing and pricing, real-time power dispatching, and ensuring the safe operation of a power grid.

In contrast to medium and long-term PV power forecasts, differentiating between short-term and ultra-short-term forecasting can not always be straightforward. Ultra-short-term PV power forecasting typically requires high-frequency data collection and short-term forecasting techniques, such as using minute or second-level data and employing machine learning, neural networks data-driven methods, and other data-driven methods. However, Short-term PV power

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed F. Zobaa .

forecasting typically involves predicting the photovoltaic power generation for the upcoming hours or days and typically relies on low-frequency data collection and longer-term forecasting methods [3]. These methods include the use of hourly or daily-level data, as well as statistical and traditional time series analysis.

In practical forecasting applications, the timescale required to achieve the desired forecasting objective frequently dictates the choice of a forecasting method. In this paper, we aim to enhance the control of power grid dispatching, enabling the power grid to accommodate fluctuations in power demand and achieve more excellent stability. To achieve this, considering the actual data sampling interval, we try to forecast PV power with the utmost accurate data granularity available through the device (15min). So this paper will focus on an in-depth investigation of ultra-short-term PV power forecasting.

Ultra-short-term PV power forecasting is accomplished via two distinct approaches: physics-based and data-driven methods [4], [5]. The physics-based approach entails establishing a physical model that describes the conversion of solar radiation into PV output power. However, this modeling process is often complex and challenging [6], [7], [8], [9], [10]. The data-driven method is a more economical and effective method, where a forecasting model is developed by analyzing historical meteorological and power data. The selection of effective samples from historical data to form the basis of the modeling process is a crucial aspect of PV power forecasting. At present, there are two main ways to optimize training samples.

The first method is “decomposition and forecasting,” that is, the original time series is decomposed and then predicted separately, and finally combined to get the final forecasting result [11], [12], [13]. Another strategy to enhance the quality of the training sample is the formation of a meteorological matrix, which describes the day’s weather conditions. Further analysis of the basic weather types, such as sunny, cloudy, and rainy, is carried out using readily available meteorological data [14], [15], [16]. To find a method that can obtain appropriate training samples under various complicated weather conditions, indicators such as cosine similarity [17], Euclidean distance [18], and gray correlation [19] are often adopted to evaluate the similarity among samples. Furthermore, historical samples whose indicators are higher than the threshold value are selected as training sets and verification sets to analyze the similarity of PV power between the forecast day and the historical sample. Yet there is a serious problem with these indicators. However, there is a significant issue with these indicators as they fail to distinguish the importance of various meteorological features in affecting PV power generation.

The time series forecasting methods of PV power are mainly based on statistical learning, machine learning, and deep learning. Statistical learning approaches have the advantage of being able to capture local patterns within data and

have been shown to effectively process non-stationary time series data [20], [21], [22]. However, these methods often require a substantial amount of historical data to be modeled, and the forecasting accuracy can be significantly impacted by the degree of data dispersion, hindering the extraction of deep features in the PV time series. Machine learning can effectively improve the above problems by adding nonlinear factors to forecasting [23], [24], [25], [26], [27], [28], yet when the data feature dimension and the number of samples increase, the complexity of feature engineering would rise sharply.

Deep learning has been a driving force in the artificial intelligence revolution due to its exceptional performance across a wide range of tasks, outperforming traditional machine learning techniques. One of the key benefits of deep learning is its ability to automatically learn hierarchical representations of data, enabling it to capture complex patterns and relationships within the input data. For instance, [29] demonstrated that combining deep convolutional neural network (CNN) architectures with Cohen’s class time-frequency representations leads to high-performance metrics in the detection of non-stationary gravitational wave signals in noisy environments. Similarly, [30] proposed a deep learning model consisting of a CNN composed of four local feature-learning blocks, achieving favorable results on several standard speech emotion datasets. Additionally, deep learning algorithms are suitable for large-scale multivariate photovoltaic power data and can self-adaptively extract features at different levels [31], [32], [33], [34].

In the realm of multivariate time series forecasting, particularly in the domain of PV power forecasting, the presence of numerous features can pose a challenge for single-class models to effectively learn and incorporate these features. To overcome this challenge, the combination of a CNN-RNN class model leverages the strengths of both models, taking advantage of the CNN’s ability to extract local features and the RNN’s ability to learn time series patterns, ultimately leading to an improvement in the performance of the forecasting model [35], [36]. But limited by the structure itself, a combined model of CNN-RNN class still has the following shortcomings:

1) CNN receptive field’s limitation: CNN extracts local features from time-series data by multiplying sliding convolution kernel Windows or average pooling. However, due to the limitation of the convolution kernel size and the convolution model, the capability of feature extraction would gradually weaken when the number of convolution operation stacks increases.

2) Disorder of feature extraction by CNN: there is no strict one-way sequence when CNN is extracting features, so the consistency of data time sequence cannot be guaranteed on the time axis.

3) RNN class model’s incapability of learning long-period patterns of time series: although variants of RNN such as gated recurrent unit (GRU) and long short-term

Memory (LSTM) can alleviate the long-term dependence problem of traditional RNN and effectively learn previously extracted local features, the issue of gradient disappearance and explosion arises when the time step is substantial, making it difficult to capture the long-term periodicity evident in photovoltaic power data over days.

Given the shortcomings of the above data-driven methods, this paper proposes a phased PV power forecasting method under complex weather, with the main contributions as follows:

1) In the meteorological analysis stage, this paper proposes a novel approach to the training sample extraction method that takes into account the importance of features in predicting PV output. This approach uses MIC to weight gray correlation analysis to extract training samples that are more similar to the forecasting samples. Compared with traditional methods that do not distinguish the importance of features, our method better considers the influence of different meteorological characteristics on PV generation. By improving the quality of training samples, our approach enhances the accuracy of ultra-short-term PV forecasting, promoting real-time dispatching of a power grid and balanced decision-making of power producers.

2) In the training model's feature extraction stage, this paper explores the use of TCN for local feature extraction of input signals. In contrast to traditional CNN serial and parallel superimposed feature extractors, TCN expands the local receptive field of convolution operation and can extract long intervals and non-continuous information of PV. Besides, the output of the feature extraction stage is consistent with the recurrent network on the time axis.

3) In the training model's forecasting stage, given the short-term regularity of the PV sequence and its periodicity with the smallest unit of a day, this paper adopts a skip connection strategy based on BiGRU to prolong the interaction cycle of time series forecasting and effectively capture the longer periodicity. Furthermore, the attention mechanism is employed to optimize the output, thereby further enhancing the forecasting performance.

II. EXTRACTING SIMILAR SAMPLES

As proposed in this paper, the overall process of the phased PV power forecasting method under complex weather is shown in FIGURE 1. Stage 1 seeks and optimizes training sets and validation sets of multiple power sequences by inputting meteorological features of historical and forecast days and according to the correlation between meteorological features and power. Stage 2 mainly builds the forecasting model, optimizes the model parameters, and evaluates the forecasting effect.

A. GREY CORRELATION ANALYSIS

Affected by meteorological factors and geographical conditions, Photovoltaic power shows strong randomness and volatility. For photovoltaic power stations with fixed locations, the size of photovoltaic power mainly depends on

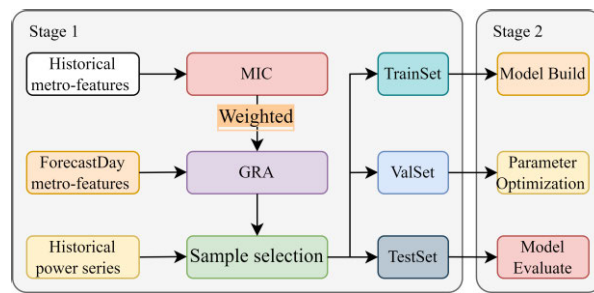


FIGURE 1. Two-stage ultra-short-term PV power forecasting process.

meteorological conditions. By analyzing the daily meteorological characteristics and applying the evaluation indexes, researchers can obtain samples whose power curve is similar to that on the forecast day, and therefore can reduce the model training time and improve the power forecasting accuracy. This paper used MIC to weight the grey correlation degree to extract samples with similar photovoltaic power, which plays a role in optimizing the training set and the validation set.

1) CONSTRUCTING METEOROLOGICAL FEATURE VECTOR

To extract similar samples, a preliminary analysis of photovoltaic output factors is necessary, so it is essential to construct a meteorological feature set that can approximate the power generation status of the day. Literature [13], [14], [15], [16], [17], [18] offers valuable insights into effective meteorological characteristics, such as average temperature, average humidity, and average irradiation intensity. Combined with the actual daily meteorological data of the photovoltaic sites, the total solar radiation data is used to displace the missing average radiation intensity data in this study. In addition, there are obvious differences among the power curves of different weather types. This paper constructs the following meteorological feature vector W for three generalized weather types, which serves as the basis for similar sample extraction:

$$W = [T_a, T_M, T_m, T_a, H_a, H_M, H_m, I_a, C_a] \quad (1)$$

where T_a , T_M , and T_m represent daily average temperature, maximum temperature, and minimum temperature respectively; H_a , H_M , and H_m represent daily relative humidity, maximum humidity, and minimum humidity respectively; I_a represents total solar radiation, and C_a represents total cloud cover.

Normalization can eliminate or reduce the influence brought by different characteristics' different units and different orders of magnitude, which is convenient for subsequent similar sample selection and photovoltaic power forecasting. After normalization, the meteorological feature vector of the forecast day and that of the same weather type are as follows:

$$W_0 = [W_0(1), \dots, W_0(n), \dots, W_0(8)] \quad (2)$$

$$W_i = [W_i(1), \dots, W_i(n), \dots, W_i(8)] \quad (3)$$

where W_0 is the meteorological feature vector of the forecast day, and $W_0(n)$ is the n^{th} meteorological feature of the forecast day; W_i represent the i^{th} meteorological feature vector of the same condition of weather, while $W_i(n)$ is the n^{th} meteorological feature of the day.

2) MIC CORRELATION ANALYSIS

MIC is a modern correlation analysis method that can fully investigate linear and nonlinear relationships between variables. Compared with the commonly used mutual information and Pierce coefficient, MIC exhibits superior equivalence and universality. For the scatter plot composed of two continuous variables with arbitrary forms and different noises, there is always a reasonable division to reflect the correlation between the variables. For the n^{th} meteorological feature set $W(n) = \{W_1(n), W_2(n), \dots, W_i(n), \dots, W_d(n)\}$, and power sequence $R = \{r_1, r_2, \dots, r_d\}$, the MIC between the n^{th} meteorological feature and the daily average power is as follows:

$$MIC(n) = \max_{a \times b < B} \frac{I(W(n); R)}{\log_2(\min(a, b))} \quad (4)$$

where $I(W(n); R)$ is the mutual information size between $W(n)$ and R , d is the number of samples under this kind of weather, a and b are the number of divided segments of $W(n)$ and R during grid division, respectively, and b , set 0.6 in this paper, is the upper limit of grid division.

3) MIC WEIGHTED GREY CORRELATION DEGREE

Grey correlation analysis can judge whether the reference data column is closely related to several comparison data columns by determining their geometric similarity, reflecting the correlation between sequences. The traditional grey correlation method calculates the correlation by average weighting, but ignores the influence of different meteorological characteristics on power. Hereby, this paper used MIC to weigh the grey correlation coefficient, and calculated the grey correlation degree between historical samples and forecast day samples under the same weather, providing a data basis for the extraction of similar samples. The specific formulae are as follows:

$$\xi_i(n) = \frac{\min_i \min_n \Delta_i + \rho \max_i \max_n \Delta_i}{\Delta_i + \rho \max_i \max_n \Delta_i} \quad (5)$$

$$a_n = \frac{MIC(n)}{\sum_{j=1}^N MIC(j)} \quad (6)$$

$$R_i = \sum_{n=1}^N a_n \xi_i(n) \quad (7)$$

where $\Delta_i = W_0(n) - W_i(n)$, ρ , the resolution coefficient was set at 0.5; a_n is the weighting coefficient of the n^{th} meteorological feature; $\xi_i(n)$ is the correlation coefficient of the n^{th} meteorological feature of the i^{th} sample under the same weather condition. R_i is the MIC-weighted grey correlation degree.

B. FORECASTING DATA SET CONSTRUCTION

1) DESCRIPTION OF THE ORIGINAL DATA SET

The multivariate power data in this paper were from the monitor data of a power station in HaiNing city from July 31, 2020 to July 31, 2021. The data set includes 20 electrical data, such as active power and equipment temperature, and eight environmental data, such as temperature, humidity, and wind speed. The sampling period was 15 minutes. In order to ensure the exclusion of redundant data, the dataset was constrained to a time range from 7:00 to 18:00 each day. A total of 45 sample points, with a 15-minute interval between them, were selected for analysis. The daily meteorological data used for similar sample extraction came from the data collected by the HaiNing station of the National Meteorological Center of China, including 8 features such as average temperature and average humidity. The data granularity was 1 group per day. The details of the data set are shown in the TABLE 1. All experimental data are normalized to eliminate the influence of different features' units and sizes on the experiment, and speed up the training of the model.

TABLE 1. Original data set.

Data type	Data name	Unit	Dimension
Multivariate power data (15min)	Active power	(kW)	1
	Equipment temperature	(°C)	1
	AC voltage	(V)	3
	AC current	(I)	3
	DC voltage	(V)	6
	DC current	(I)	6
	Grid frequency	(HZ)	1
	Wind speed	(m/s)	1
	Wind direction	(c)	1
	Environment temperature	(°C)	1
Meteorologic al Date(day)	irradiation intensity	(W/m2)	2
	irradiation amount	(MJ/m2)	2
	(Max/Min/Ave) temperature	(°C)	3
	(Max/Min/Ave) humidity	%	3
	Total amount of irradiation	(MJ/m2)	1
	Total cloud cover	%	1
	The weather types	/	1

2) CONSTRUCTION AND DIVISION OF DATA SETS

The sliding window method was used to construct the sample set for training and testing models. To verify the validity of the forecasting and stability of the model, forecastings were conducted after 15 minutes, 30 minutes, 45 minutes, and 60 minutes, respectively. The process of constructing the sample set is shown in FIGURE 2.

The power data and feature data within the sliding window size T which is the length of the input vectors at the current, and be used to predict the power size of multiple time lengths. In addition, similar samples with a total of 25 days were obtained via screening the selected test samples on the forecast day with MIC-optimized gray correlation analysis. They were arranged in chronological order, and the training set and the verification set were divided according to the proportion

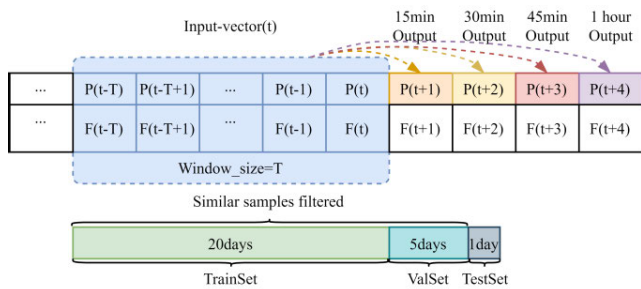


FIGURE 2. Four different time scale forecasting methods and dataset partitioning process.

in FIGURE 2. In particular, the training set was used to build the basic model for power forecasting, the verification set was used to adjust parameters and optimize the model, and the test set was used to evaluate the performance of ultra-short-term power forecasting. Given the intricate nature of PV power time series and the end-to-end learning mechanism of the neural network, the preservation of the model’s robustness necessitated the avoidance of manual construction of high-order features through feature engineering methods. This approach was aimed at preserving the full effectiveness of the original sequence by minimizing any alteration or distortion to the available information.

III. FORECASTING MODEL CONSTRUCTION

The structure of the overall forecasting model (TCN-SK-BiGRU) is shown in FIGURE 3. The input characteristics of the forecasting model at the current time were expressed as $X = \{x_1, x_2, \dots, x_t, \dots, x_T\}$, where x_t is the state matrix $[p(t), F(t)]$ of the t^{th} time step. The original input features are extracted by the TCN network layer, and then the extracted features are input to the cyclic jump network for time series forecasting.

A. TCN FEATURE EXTRACTION LAYER

1) EXTENDED CAUSAL CONVOLUTION

As a core part of TCN, extended causal convolution fully combines the advantages of both causal convolution and extended convolution. The extended causal convolution process with an expansion coefficient $d = [1], [2], [4]$ and a convolution kernel $K = 3$ is shown in FIGURE 4. In particular, causal convolution strictly controls the sequence of convolution operations to ensure that the outputs will not omit any historical information. Besides, by gradually increasing the expansion coefficient d or convolution kernel K , expansion convolution changes the size of the receptive field $R = d * K$, reduces the number of layers of the network, and increases long-term memory. Additionally, the input and the output of a TCN become equal in length by the same convolution operation and zero-value filling. For the extended causal convolution’s input $X^{(n-1)} = \{x_1^{n-1}, \dots, x_t^{n-1}, \dots, x_T^{n-1}\}$ and its convolution kernel $K = \{k_1, k_2, \dots, k_m\}$, the output of the t^{th} unit $o_t^{(n)}$ after the extended causal convolution is calculated in

Formula (8):

$$o_t^{(n)} = (X^{(n-1)} * dK)(s) = \sum_{i=1}^m k_i \cdot x_{t-d \cdot i}^{(n-1)} \quad (8)$$

where $*$ is convolution operation, and m is the number of convolution kernels.

2) TCN UNIT

A simple serial superposition of extended convolution can further enlarge a receptive field. However, when a network gets deeper, the network will overlearn the parameters of non-identity mapping as the convolution superposes, resulting in information loss during the training process. Therefore, as shown in FIGURE 5, a residual connection is added between TCN units, and the input data passes through the additional channels to transfer information across hierarchies, which, in the meantime of superposing and expanding the causal convolution, to some extent solves the problems of network degradation and gradient fragmentation caused by the increase of neural network depth. Wherein, 1×1 convolution ensures the same dimension of residual connection. The residual connection process between TCN units is as follows:

$$X^{(n)} = \text{Activation} \left(X^{(n-1)} + f_S \left(X^{(n-1)} \right) \right) \quad (9)$$

where f_S represents the twice convolution, normalization, ReLU and Dropout operations within each TCN unit.

To achieve faster convergence speed and stronger learning rate robustness during network training, the weight vector is re-parameterized by weight normalization after each extended causal convolution, divided into modular value m and direction V . As shown in Formula (10), weight normalization is more suitable for analysis of time-series problems.

$$W = \frac{m}{\|V\|} V \quad (10)$$

Finally, the normalized results are input into the ReLU activation function to add nonlinear factors to the network, and a Dropout is performed to randomly discard some hidden layer neurons to simplify the overall parameters of the network and avoid over-fitting.

3) TCN-BASED FEATURE EXTRACTION

As a variant of CNN, TCN realizes the function of sparse multiplication and weight sharing, fully mining local features of data and meanwhile reducing parameters. In addition, the direction of back propagation is different from that of the time sequence during model training, which avoids the accumulation of gradient in the time dimension. Based on CNN, TCN uses extended convolution to stack convolution layers with different expansion coefficients, expanding feature extraction’s reception field in the time dimension while keeping the network parameters unchanged. Causal convolution can maintain the sequential nature of photovoltaic power data, that is, after the data of each time step is processed by the causal convolution layer, the direction of the output and that of the cyclic network are kept the same on the time axis.

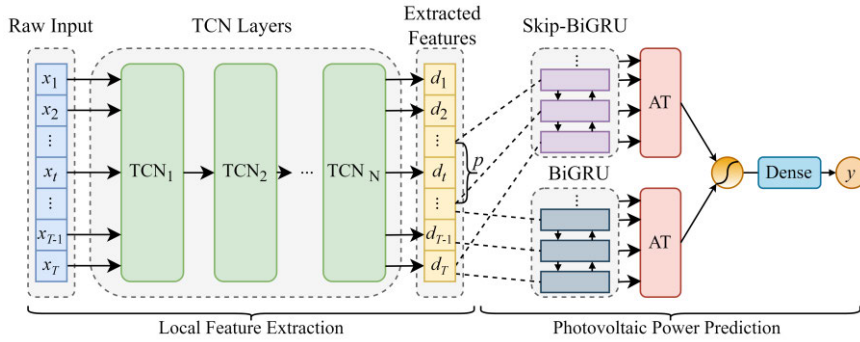


FIGURE 3. The overall structure of the proposed forecasting model.

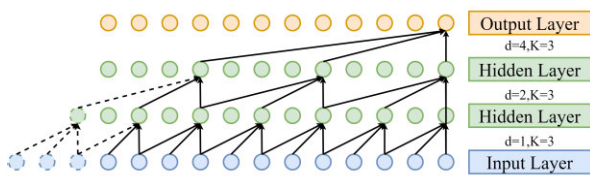


FIGURE 4. Extended causal convolution.

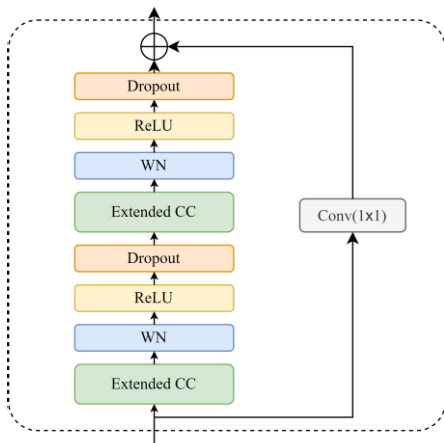


FIGURE 5. Structure diagram of TCN residual unit.

In this paper, the TCN structure was used as the feature extractor of the photovoltaic power forecasting model with a large number of features. For the original input feature $X = \{x_1, x_2, \dots, x_t, \dots, x_T\}$, the output of the feature extraction layer through a residual connection of n TCN units is $D = \{d_1, d_2, \dots, d_t, \dots, d_T\}$, where $d_t = x_t^{(n)}$. The feature extraction results need to be input into the subsequent cyclic jump network to further extract the patterns of time series.

B. SKIP CONNECTION LAYER

1) BiGRU

Since the transmission orders of sequential information in both causal convolution and recurrent neural networks are unidirectional, GRU only considers the unidirectional data

information flow and ignores the influence of reverse data sequences' transformation law on short-time power forecasting. BiGRU can further learn the time series characteristics but has not been fully used in the photovoltaic field. In this paper, BiGRU was used as the basic model to capture temporal patterns, as shown in FIGURE 6, including forward-GRU, backward-GRU, and Concatenate processes.

The unit feature d_t , after extraction by the TCN layer, of the t^{th} hidden layer is affected by the state before and after adjacent time. Using the BiGRU network to calculate the forward propagation state and the back-propagation state respectively, then concatenate the hidden layer results of the two propagation states based on channel dimensions to obtain the bidirectional time-series feature. The calculation process is as follows:

$$\vec{s}_t = \overrightarrow{GRU}(d_t, \vec{s}_{t-1}, \vec{c}_{t-1}) \tag{11}$$

$$\overleftarrow{s}_t = \overleftarrow{GRU}(d_t, \overleftarrow{s}_{t+1}, \overleftarrow{c}_{t+1}) \tag{12}$$

$$s_t = [\vec{s}_t; \overleftarrow{s}_t] \tag{13}$$

where \vec{s}_t is the forward propagation state of unit t in the forward GRU layer, \vec{c}_{t-1} is the cell state of unit $t-1$ in the forward GRU layer, \overleftarrow{s}_t is the forward propagation state of unit t in the reverse GRU layer, \overleftarrow{c}_{t+1} is the cell state of unit $t+1$ in the reverse GRU layer, s_t is the output unit of

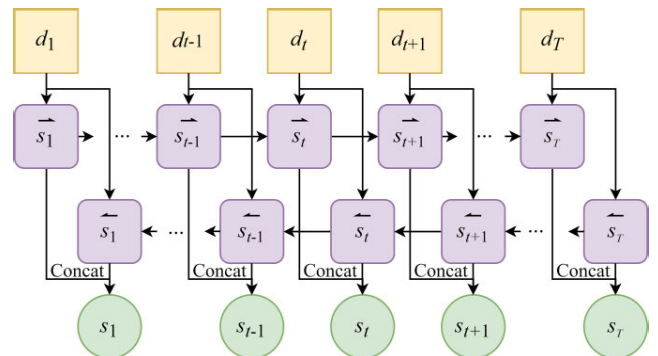


FIGURE 6. BiGRU structure diagram.

unit T in BiGRU, and $[\cdot]$ is concatenate of matrix based on the channel dimension.

2) CIRCULAR ATTENTION MECHANISM

One can obtain the power state at each moment by learning temporal patterns via BiGRU. The historical power state will affect the current power, and the influence of near states and the influence of far states on the power size are different, generally showing that a near state is larger than a far one. When the final hidden layer unit is utilized as the output straightly, the valuable information contained in other hidden layer units remains untapped, ultimately compromising the model's accuracy as the time step and the number of hidden layer units increase. According to the influence of input data on outputs, the sequential attention mechanism assigns different weights to hidden layer states, and, thus, fully evaluates the importance of each hidden layer unit, namely the historical state of each moment. The process is as follows:

1) Equally replace the output result of BiGRU network with $H = \{h_1, h_2, h_t, \dots, h_T\}$, $t \in [1, T]$. In this paper, the additive model was used as an attention-scoring function to calculate the correlation $Score(h_t)$ between the state of the hidden layer \bar{h} and the output state at different historical moments.

$$Score(h_t) = \tanh(\mathbf{W}_s h_t + \mathbf{b}_s) \quad (14)$$

where \mathbf{W}_s is the weight matrix, and \mathbf{b}_s is the offset term.

2) Softmax function is used to calculate the attention weight a_t of the hidden layer unit h_t , that is, the proportion of attention at this moment. The formula can be expressed as follows:

$$a_t = \frac{\exp(Score(h_t, h_T))}{\sum_{j=1}^T \exp(Score(h_j, h_T))}; \quad \sum_{t=1}^T \exp(a_t) = 1 \quad (15)$$

3) Each hidden layer unit is given the corresponding attention weight, and the weighted average is used to obtain the optimized output $h = \sum_{t=1}^T a_t h_t$, of the attention mechanism. The outputs cover all the historical states within the time step.

3) SKIP CONNECTION STRATEGY

To capture the long-term pattern in the unit of days, the sequential period p is introduced as the number of jump steps. As a result, it expands the span of the time series, and changes the interaction cycle between the current forecasting time and the historical time, and that between the current forecasting time and the future time. Single point forecasting combines the jump period p and the time step Q of the cyclic network and at the moment the time window's size $T = p * Q$. The forecasting task is set to predict the power size of samples the next time according to the sample information at the first T time. The jump period p , set to 45 in this paper, is the number of samples within one day.

The state update of the hidden layer in the BiGRU network with skip connection at each moment no longer

depends on the state of the single hidden layer before or after but establishes a connection with the hidden layer with an adjacent number of jump steps to learn the long-term correlation in days in photovoltaic power forecasting. The calculation process of the Skip-BiGRU network is shown in Equations (13)~(16).

$$\vec{b}_t = \overrightarrow{GRU}(d_t, \vec{b}_{t-p}, \vec{c}_{t-p}) \quad (16)$$

$$\overleftarrow{b}_t = \overleftarrow{GRU}(d_t, \overleftarrow{b}_{t+p}, \overleftarrow{c}_{t+p}) \quad (17)$$

$$b_t = [\vec{b}_t; \overleftarrow{b}_t] \quad (18)$$

where \vec{b}_t is the forward propagation state of the t^{th} unit in the forward GRU layer, \vec{c}_{t-p} is the cell state of the $t-p^{th}$ unit in the forward GRU layer, \overleftarrow{b}_t is the forward propagation state of the t^{th} unit in the reverse GRU layer, \overleftarrow{c}_{t+p} is the cell state of the $t+p^{th}$ unit in the reverse GRU layer, b_t and is the output of the t^{th} unit of BiGRU that adopts skip connection strategy.

The output hidden layer S of BiGRU and the output hidden layer B of Skip-GRU are optimized respectively through the Attention layer, and then concatenate the optimization results, which are input into a fully connected network with a Sigmoid activation function. Consequently, the final forecasting result y of the whole model can be obtained.

$$B = \{b_1, b_p, \dots, b_{pi}, \dots, b_{pQ}\} \quad (19)$$

$$S = \{s_{(p-1)Q+1}, s_{(p-1)Q+2}, \dots, s_{(p-1)Q+i}, \dots, s_{pQ}\} \quad (20)$$

$$A = \text{Concat}[\text{Attention}(S), \text{Attention}(B)] \quad (21)$$

$$y = \text{sigmoid}(W_f A + b_f) \quad (22)$$

where \mathbf{W}_f is the weight matrix, \mathbf{b}_f is the bias term, and $i \in [1, T]$.

IV. EXPERIMENTAL ANALYSIS

On sunny, cloudy, and rainy days, and three-day power data were used as test sets respectively in this paper. Specific training sets and validation sets were searched and optimized according to the one-day weather conditions of the test sets, and corresponding forecasting models were built to predict ultra-short-term power in the four time-scales.

A. EXTRACTION RESULTS OF SIMILAR SAMPLES

MIC analysis was conducted on the daily meteorological data within one year under the three weather types, and the analysis results are shown in TABLE 2. Based on our analysis, it is evident that the size of the output of PV power is influenced by varying meteorological characteristics, with discernible differences observed between different types of weather. Notably, the total radiation exhibited the strongest correlation with power output across all three weather types, with MIC values exceeding 0.92. This meteorological feature was found to be significantly more impactful on PV power output than other weather parameters. On sunny days, due to stable weather conditions, the correlation between power and temperature was the highest. Under cloudy conditions, there were high correlations between power generation and cloud

TABLE 2. Analysis results based on MIC.

Type	Highest temperature	Lowest temperature	Mean temperature	Maximum humidity	Lowest humidity	Average humidity	Total cloud volume	Total radiation
Sunny	0.563	0.543	0.595	0.296	0.322	0.427	0.245	0.967
Cloudy	0.427	0.403	0.450	0.348	0.512	0.585	0.541	0.929
Rainy	0.373	0.294	0.374	0.349	0.607	0.625	0.290	0.926

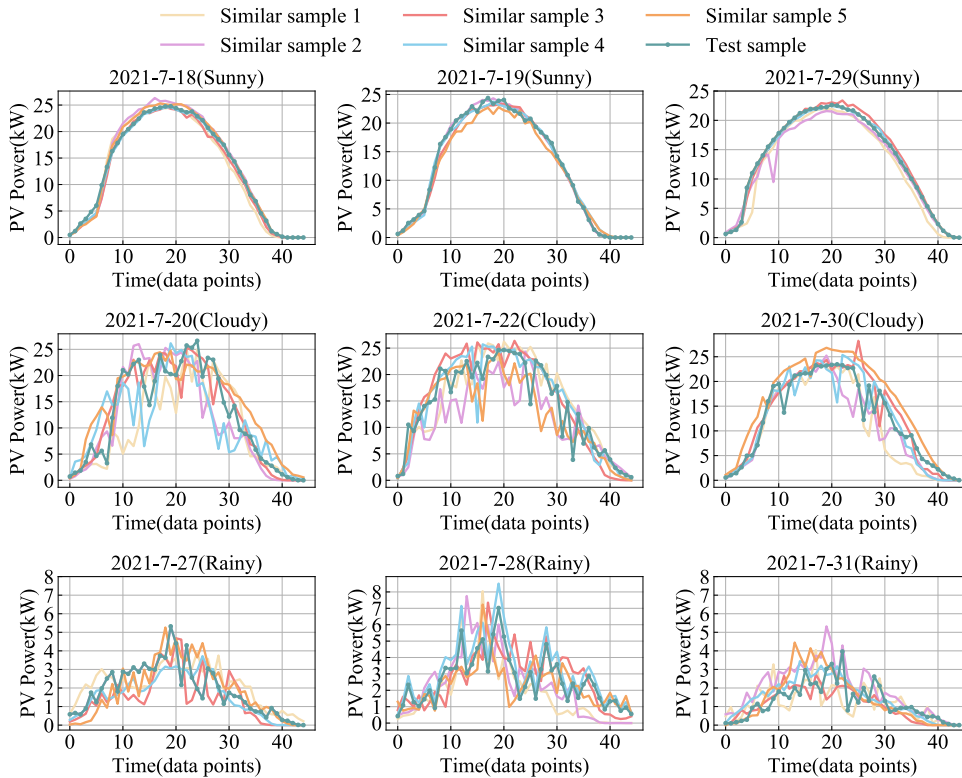


FIGURE 7. Similar samples under three weather types (sunny, cloudy, rainy) for the specific test samples identified extracted by MIC weighted grey correlation analysis.

cover, temperature, and humidity. When it was cloudy and rainy, the weather conditions were complex, the moisture and air changed obviously, and the power was largely determined by the humidity characteristics among which the average humidity has the most significant influence with the MIC reaching 0.625.

Similar samples were extracted by MIC weighted grey correlation analysis for the specific test samples identified. The extraction results are shown in FIGURE 7. To ascertain the efficacy of analogous sample extraction, we exhibited the power curves of the top five days (comprising five akin samples) with the highest weighted grey correlation for every forecasting day. Evidently, on sunny days, the sample degrees were exceptionally elevated, which ensured a persistent state of stable power generation. In contrast, the analogous samples collected during cloudy weather showed some fluctuations; however, they could still attain their maximum potential around noon, with stable extreme points ranging from 20 to 25kW. The training on overcast and rainy days had a great fluctuation, but the changing trends of the curves were

roughly the same. The above screening of similar samples can eliminate redundant power data, improve the quality of samples, reduce the training time of the subsequent forecasting model, and improve the forecasting accuracy of ultra-short-term power.

B. EVALUATION METRICS

Mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R^2) were used to evaluate the forecasting accuracy of the experimental results in this paper. The calculation formula for each evaluation index is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{23}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \tag{24}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - y_{mean})^2} \tag{25}$$

where n is the number of predicted samples, $\hat{y}_i = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$ is the predicted value of samples, and $y_i = \{y_1, y_2, \dots, y_n\}$ is the actual value of samples, and $y_{mean} = \frac{1}{n} \sum_{i=1}^n y_i$.

Besides, to prove the robustness of the model, model construction and power forecasting would be carried out under three weather conditions and three forecasting scales respectively. Furthermore, we will introduce the model construction time and forecasting time to undertake a more comprehensive evaluation of the promptness of the predicted outcomes.

C. MODEL PARAMETER SETTING

Given the experiment’s characteristic of having a moderate data scale and several parameters, a limited range was preset for each parameter, and the optimal results of the experimental parameters and those of the training parameters were searched by the grid search method, where the channel number = {4, 8, 16, 32, 64 }, dropout = {0.1, 0.2, 0.3}, cell number = {1, 2, 3, 4}, TCN expansion coefficient = {1, 2, 4, 8, 16, 32}, TCN convolution kernels size = {1, 2, 3, 4, 5}, and batch_size = {4, 8, 16, 32}. TABLE 3 shows the parameter configuration optimized by grid search.

TABLE 3. Optimized parameter table.

Source	Parameter types	Parameters size
TCN layer	channel number	64
	Unit number	2
	Dropout	0.2
	Expansion coefficient	[1,2,4,8]
	Kernel size	2
	Activation function	ReLU
BiGRU layer	Channel number	32
	Layer number	2
	Dropout	0.3
	Activation function	Sigmoid
Training parameters	Batch size	16
	Optimizer	Adam
	Loss function	MSE
	Number of iterations	200
	Initial learning rate	0.01

To ensure that the model was fully trained, the method of learning rate attenuation was adopted to gradually learn from shallow to subtle. The initial learning rate was 0.01, the lowest learning rate was 0.0001, the attenuation coefficient was 0.5, and the patience was 5. Since the model adopts sliding window to predict photovoltaic power, the sliding window size T has a great difference in forecasting performance under different weather conditions. To observe the forecasting effect of different time window sizes, centralized tuning of time window sizes is required. Since $T = p * Q$, the time step Q of the circular network was searched while keeping the other parameters unchanged and setting the time step to 1, 4, 8, 12, 16, and 20. Taking the mean of R^2 as evaluation matrix, the forecasting results under three weather type are shown in FIGURE 8. The analysis shows that the optimal forecasting

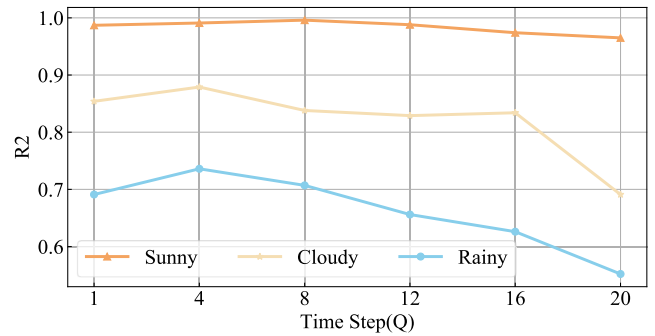


FIGURE 8. Use R^2 as the evaluation metric, set the time step 1, 4, 8, 12, 16, 20. the forecasting results under three weather type.

effect was achieved when the time window size was set to 8, 4, and 4 respectively under sunny, cloudy, and rainy weather.

D. COMPARATIVE EXPERIMENTAL

To evaluate the predictive performance of the proposed TCN-BILSTM model (SK-TCN-BILSTM) with skip connection strategy, we constructed datasets using MIC weighted gray correlation degree for all models prior to forecasting. The data sets were created with identical time window sizes and proportions to ensure a uniform experimental environment. To benchmark the proposed model, we compared it with various statistical models such as ARIMAX, exponential smoothing (ES), and GPR [19], which are well-suited for large datasets, as well as machine learning-based models including XGBOOST [26], SVM[100], and deep learning-based models such as CNN [28], LSTM [29], GRU [30], LSTM-Attention(LSTM-A) [31], CNN-LSTM [32], and ConvLSTM [33]. The parameters of all algorithms in the comparison test are optimized for this situation, and the significant hyperparameter settings of all the models mentioned above are shown in TABLE 4.

The experimental results are shown in the following table. To ensure the stability of the experimental results, the final results of the following experiments at each moment were the average values after five times of model training and testing, and the results of forecasting are shown in TABLE 5-8.

In general, the experiment conducted model training and predictive analysis on four time-scales of 15, 30, 45, and 60 minutes respectively to verify the robustness of the model. When the forecasting time increased, we put forward higher requirements on the learning ability and data analysis ability of the model. For sunny weather, the forecasting accuracy of the four-time scales was very close, with a small decrease after 60 minutes. For cloudy and rainy weather, it can be observed that the forecasting error of the model increased significantly in longer time scales.

Specifically, the power data had strong stability on sunny days, and the forecasting effects of all models were considerable. Machine learning-based models and statistical models also performed well on sunny days and were even better

TABLE 4. The significant hyper-parameter settings of all the models.

Model types	Model	Model	Parameters
Statistical models	ARIMAX	Order	(1,0,1)
		Exog	Feature matrix
	ES	Trend	Add
		Seasonal	Add
	GPR	Seasonal periods	45
Kernel		Rbf	
optimizer number		10	
Machine Learning-based models	SVR	Alpha	0.1
		Kernel	Rbf
		Gamma	0.01
		Degree	3
		Epsilon	0.1
	XGBoost	Cache size	200
		C	10
		Booster	Gbtree
		Objective	Logistic reg
		Max depth	11
Deep Learning-based models	CNN	Eval metric	MAE
		Eta	0.05
		Verbosity	0
		Unit number	3
		Filters	32 / 32 / 32
	GRU/ LSTM	Stride	4 / 3 / 2
		Padding	Same
		Activation function	ReLU
		Layer number	2
		Channel number	64 / 64
Combined models	ConvLSTM	Dropout	0.3 / 0.1
		Activation function	Sigmoid
		Layer number	3
		Filters	20 / 10
		Kernel size	(1,1)/(3,3)/(1,1)
	LSTM-A	Padding	Same
		Activation function	Sigmoid
		LSTM parameter	Same as LSTM
		Scoring function	Additive
		CNN-LSTM	CNN parameter
LSTM parameter	Same as LSTM		

than the single CNN and RNN class models in multiple time scales. However, in cloudy and rainy weather with frequent meteorological fluctuations, the methods based on statistics and machine learning could not fully analyze and learn the high-dimensional features, and the forecasting accuracy would decrease sharply, especially for ARIMAX. At this point, the method based on deep learning showed obvious advantages. In particular, CNN automatically extracted deep features, which was better than the traditional method with three evaluation metrics, but could not capture the historical patterns of power. For the RNN models, compared with CNN, which is suitable for time series analysis by connecting the front and rear units, LSTM and GRU had very similar forecasting performance in the three types of weather, but the forecasting accuracy of a single model was still low. Compared to the single LSTM, the addition of Attention improved the forecasting performance in each time scale and under all weather conditions.

The amalgamated CNN and LSTM model harnesses the respective strengths of both architectures, enabling comprehensive exploration of local features and time-series

patterns. This synergistic approach results in superior forecasting accuracy when compared to individual models operating in isolation. Compared with the single CNN class and RNN class models, the CNN-LSTM and the ConvLSTM both had somewhat higher precision under sunny conditions, with R2 above 0.99 in all four-time scales. In particular, in predicting an hour later, the R2 of CNN-LSTM is the best, 0.995, while the RMSE and MAE of ConvLSTM are the lowest which are 0.630 and 0.476. In cloudy and rainy weather, CNN-LSTM and ConvLSTM had similar performances. Specifically, ConvLSTM was slightly better than CNN-LSTM in predicting 15 minutes and 30 minutes later, and CNN-LSTM had better performance in predicting 45 minutes and an hour later.

The method proposed in this paper extracted features through the TCN and adopted a jumping strategy to learn time-series patterns. The method not only ameliorated the unsoundness and limitations of feature extraction of the CNN but also solved LSTM's and GRU's incapability of fully learning long periodicity. Generating the forecasting results better than the other benchmark models in all the time scales and under all the weather types, the method showed strong abilities in learning and nonlinear training. On sunny days with obvious periodic regularity, the proposed method performed best in the treble scales, while its performance was slightly worse than CNN-LSTM and ConvLSTM in predicting an hour later. In cloudy and rainy weather, where feature extraction was difficult, the ability of the model to analyze data became particularly important. Compared with the combined model of CNN-LSTM and ConvLSTM, the proposed method had R² improved by 2.5% and 2.8% on average, RMSE decreased by 7.1% and 6.3%, and MAE decreased by 6.8% and 7.8% under cloudy conditions. In rainy weather, the forecasting performance of the proposed method had the greatest improvement. Particularly, when the time interval increased, the forecasting effect of the other models plummeted, but the method can still maintain R² at 0.7369 in predicting after 45 minutes. However, by observing the forecasting result an hour later, it can be found that, due to complex environmental changes, it was still difficult even for our model to predict generating capacity accurately. Despite our best efforts, we found it challenging to accurately predict generating capacity using our model, even after a one-hour interval, due to the intricate and dynamic nature of the environment. However, we established that the TCN-SK-BIGRU model is highly appropriate for power forecasting under diverse weather conditions, and possesses exceptional self-adaptability that enables ultra-short-term forecasting across most time intervals. Notably, this model outperforms other combined models in terms of stability and generalization, underscoring its efficacy as a robust and versatile forecasting tool.

Finally, by analyzing the training time and the forecasting time of the model, it is not hard to find, under the same weather conditions, the same class model's training time and testing time in the different time scales basically remained

TABLE 5. R², RMSE, MAE, training Time(Time1), forecasting Time(Time2) when time scale is 15 minutes.

Data	Sunny					Cloudy					Rainy				
	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2
ARIMAX	0.984	1.187	0.915	15.273	0.017	0.847	3.507	2.763	16.012	0.019	0.657	0.721	0.592	17.538	0.018
ES	0.983	1.287	1.031	0.445	0.009	0.826	3.694	2.883	0.433	0.009	0.641	0.728	0.618	0.567	0.010
GPR	0.983	1.246	1.046	13.958	0.068	0.843	3.583	2.736	12.580	0.042	0.664	0.736	0.578	16.375	0.048
SVR	0.984	1.208	1.007	1.249	0.009	0.842	3.596	2.773	1.295	0.010	0.667	0.732	0.593	1.279	0.009
XGBoost	0.985	1.173	0.962	13.347	0.005	0.849	3.505	2.754	15.002	0.006	0.675	0.727	0.588	16.082	0.005
CNN	0.991	0.845	0.702	39.763	0.057	0.882	3.082	2.226	47.283	0.089	0.719	0.682	0.462	50.230	0.036
LSTM	0.988	1.092	0.899	71.832	0.102	0.898	2.861	2.177	86.928	0.127	0.703	0.702	0.521	106.987	0.110
LSTM-A	0.990	0.931	0.732	83.569	0.113	0.902	2.711	2.096	103.485	0.130	0.710	0.688	0.492	121.753	0.122
GRU	0.986	1.128	0.945	68.287	0.101	0.896	2.890	2.207	81.428	0.123	0.708	0.697	0.515	101.229	0.107
CNN-LSTM	0.993	0.758	0.611	70.854	0.132	0.909	2.587	2.003	83.493	0.129	0.735	0.663	0.476	104.944	0.122
ConvLSTM	0.994	0.707	0.576	318.234	0.327	0.912	2.599	1.986	334.952	0.298	0.742	0.657	0.465	391.301	0.401
Proposed	0.996	0.673	0.522	245.967	0.587	0.928	2.383	1.870	258.589	0.645	0.786	0.596	0.428	273.863	0.665

TABLE 6. R², RMSE, MAE, training Time(Time1), forecasting Time(Time2) when time scale is 30 minutes.

Data	Sunny					Cloudy					Rainy				
	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2
ARIMAX	0.981	1.312	1.078	15.397	0.018	0.785	4.140	3.220	17.902	0.019	0.558	0.864	0.672	18.834	0.019
ES	0.976	1.389	1.209	0.473	0.008	0.801	3.962	3.085	0.457	0.010	0.609	0.796	0.613	0.598	0.009
GPR	0.979	1.376	1.096	12.152	0.068	0.821	3.773	2.946	13.489	0.050	0.646	0.767	0.573	15.507	0.035
SVR	0.983	1.257	1.057	1.332	0.010	0.828	3.789	2.813	1.307	0.010	0.643	0.769	0.583	1.359	0.009
XGBoost	0.986	0.879	0.956	12.234	0.005	0.826	3.724	2.742	14.882	0.009	0.660	0.751	0.551	15.185	0.003
CNN	0.992	0.831	0.678	38.756	0.058	0.836	3.610	2.707	53.728	0.079	0.697	0.708	0.517	50.258	0.059
LSTM	0.982	1.266	0.986	67.395	0.106	0.845	3.511	2.577	84.875	0.101	0.675	0.734	0.510	104.134	0.104
LSTM-A	0.984	1.212	0.921	85.159	0.119	0.847	3.492	2.553	100.221	0.110	0.689	0.719	0.507	122.982	0.122
GRU	0.987	1.086	0.787	69.948	0.106	0.846	3.504	2.576	77.676	0.092	0.676	0.733	0.533	94.631	0.105
CNN-LSTM	0.994	0.748	0.560	73.142	0.120	0.852	3.429	2.528	79.816	0.120	0.714	0.689	0.476	102.985	0.124
ConvLSTM	0.990	0.924	0.695	330.536	0.290	0.854	3.412	2.526	324.255	0.312	0.725	0.675	0.466	374.190	0.582
Proposed	0.995	0.647	0.544	233.333	0.615	0.876	3.265	2.405	261.217	0.631	0.745	0.650	0.464	288.730	0.550

TABLE 7. R², RMSE, MAE, training Time(Time1), forecasting Time(Time2) when time scale is 45 minutes.

Data	Sunny					Cloudy					Rainy				
	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2
ARIMAX	0.978	1.441	1.267	16.461	0.019	0.670	5.133	3.687	17.023	0.020	0.401	0.997	0.813	19.278	0.017
ES	0.973	1.587	1.131	0.446	0.010	0.723	4.567	3.408	0.469	0.009	0.513	0.912	0.711	0.513	0.011
GPR	0.982	1.284	1.089	11.759	0.063	0.786	4.134	3.262	11.981	0.033	0.562	0.852	0.645	15.961	0.058
SVR	0.985	1.161	0.994	1.308	0.009	0.766	4.322	3.318	1.271	0.011	0.611	0.804	0.607	1.411	0.010
XGBoost	0.986	1.127	0.988	10.147	0.005	0.793	4.062	3.174	15.527	0.005	0.616	0.798	0.610	16.114	0.005
CNN	0.993	0.763	0.614	40.328	0.060	0.809	3.903	2.926	46.494	0.058	0.665	0.745	0.511	51.603	0.038
LSTM	0.984	1.189	0.935	72.505	0.103	0.802	3.976	2.916	94.508	0.110	0.633	0.780	0.582	107.185	0.111
LSTM-A	0.986	1.129	0.868	81.944	0.119	0.808	3.912	2.776	106.666	0.124	0.654	0.757	0.562	117.969	0.130
GRU	0.984	1.201	0.983	69.164	0.103	0.802	3.975	2.814	86.695	0.110	0.642	0.771	0.565	102.793	0.112
CNN-LSTM	0.994	0.720	0.572	74.635	0.145	0.826	3.726	2.701	82.107	0.119	0.716	0.687	0.476	101.485	0.083
ConvLSTM	0.994	0.715	0.529	300.731	0.284	0.819	3.795	2.937	353.835	0.259	0.713	0.690	0.525	369.488	0.576
Proposed	0.996	0.574	0.491	234.248	0.468	0.850	3.459	2.478	251.668	0.682	0.736	0.658	0.473	270.769	0.867

TABLE 8. R², RMSE, MAE, training Time(Time1), forecasting Time(Time2) when time scale is 60 minutes.

Data	Sunny					Cloudy					Rainy				
	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2	R ²	RMSE	MAE	Time1	Time2
ARIMAX	0.967	1.755	1.427	16.042	0.018	0.569	5.867	4.137	17.191	0.017	0.301	1.265	1.001	20.258	0.017
ES	0.970	1.699	1.401	0.466	0.010	0.681	4.972	3.589	0.443	0.011	0.417	0.986	0.803	0.522	0.013
GPR	0.983	1.227	1.044	13.887	0.069	0.775	4.241	3.476	12.157	0.033	0.474	0.935	0.766	16.275	0.048
SVR	0.986	1.111	0.860	1.349	0.009	0.774	4.245	3.425	1.287	0.009	0.539	0.875	0.743	1.297	0.008
XGBoost	0.987	1.100	0.989	15.772	0.005	0.778	4.209	3.524	15.151	0.004	0.595	0.820	0.699	16.578	0.004
CNN	0.992	0.861	0.652	36.386	0.059	0.769	4.293	3.503	43.602	0.105	0.577	0.838	0.673	58.285	0.049
LSTM	0.981	1.325	1.075	73.642	0.101	0.782	4.175	3.245	89.524	0.113	0.561	0.853	0.672	101.225	0.105
LSTM-A	0.989	0.971	0.778	82.003	0.119	0.787	4.118	3.174	108.801	0.125	0.594	0.821	0.612	114.516	0.107
GRU	0.981	1.316	1.129	68.077	0.105	0.783	4.159	3.261	83.521	0.112	0.545	0.869	0.689	105.428	0.108
CNN-LSTM	0.993	0.636	0.500	69.348	0.124	0.806	3.930	3.056	84.853	0.121	0.656	0.755	0.563	105.862	0.122
ConvLSTM	0.993	0.630	0.476	330.423	0.372	0.800	3.994	3.172	347.767	0.273	0.649	0.763	0.551	365.125	0.519
Proposed	0.994	0.698	0.571	246.115	0.604	0.828	3.700	2.840	264.886	0.662	0.679	0.730	0.480	286.571	0.745

unchanged. Yet in different weather, the time spent on a model was directly proportional to the power curve’s miscellaneous degree and the difficulty of data analysis. In specific, the spent time was the shortest when sunny, moderate when cloudy, and the longest when rainy. In detail, both the statistics-based and machine-learning-based models require a relatively shorter time for training and prediction. Notably, the training time for ES is the briefest, approximately 0.5 seconds, while the prediction time for XGBoost is the shortest, at 5×10^{-3} seconds. The methods based on deep learning greatly improved in model training, among which ConvLSTM cost the most time fluctuating around five minutes, and the test time of each model was within 1s. The method proposed in this paper has high complexity, because it stacks TCN for local feature extraction, and uses long and short periods to learn time-series patterns. The average training time was 246.1s on sunny days, 259.1s on cloudy days, and 279.9s on rainy days, while the average test time was 0.56s on sunny days, 0.65s on cloudy days, and 0.71s on rainy days. Both training time and test time increased with the complexity of the weather. Although there is no outstanding advantage in time performance, the test time within 1s and training time within five minutes guarantee that our method can still be effectively applied in practical forecasting engineering.

E. ANALYSIS OF ABLATION EXPERIMENT

To evaluate the efficacy of each module in our forecasting methodology, we conducted an ablation experiment that involved scrutinizing the specific forecasting approaches for the three distinct weather conditions. The experiment was designed to analyze the predictive performance after a 15-minute interval. Considering the validity and rigor of the experiment, we ensured that the variable was unique, only deleted or replaced a single module, and kept the internal parameters of the replacement module consistent. The experiment names and specific settings were as follows:

- 1) NoMIC: the grey correlation degree was not used to extract similar samples of forecasting samples, and the samples under the same weather were used to divide the training set and verification set in a 5:1 ratio.
- 2) NoSKIP: the circular jump strategy was not adopted, while the forecasting was only made through the ordinary TCN-BigRU model.
- 3) NoAT: optimize the output without Attention.
- 4) NewCNN: the convolution kernel size was set to 2, the number of layers was set to 4, and the padding mode with Same replaced TCN for feature extraction.
- 5) NewGRU: BiLSTM, the base model of the cyclic Skip connection layer, was replaced with LSTM with 2 layers and 64 channels.
- 6) Proposed: the complete model proposed in this paper.

Experimental results under three weather types and three different evaluation metrics are shown in FIGURE 9. The proposed model shows the best performance under three weather types and indicators. Similar sample extraction using grey correlation degree and replacing traditional CNN local

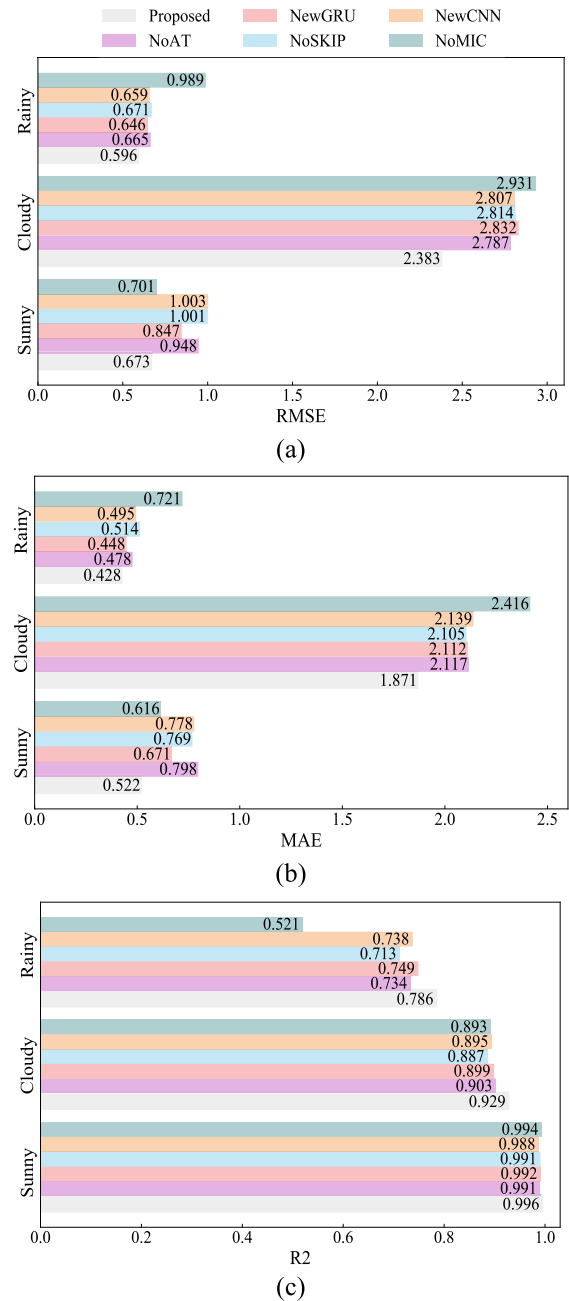


FIGURE 9. Experimental results under three weather types and three different evaluation metrics. (a) RMSE. (b) MAE. (c) R2.

parallel feature extraction with TCN had similar promoting effects. In specific, they both significantly improved the forecasting effect of the model in rainy weather with large fluctuations and slightly improved that in sunny and cloudy weather. Thus, the above two methods were important in power forecasting.

The cyclic skip connection strategy could effectively reduce the model error under sunny and cloudy conditions with obvious long periods, and could also act this way on rainy days, which proved the effectiveness of the cyclic link strategy in capturing long-term dependence. In addition, with

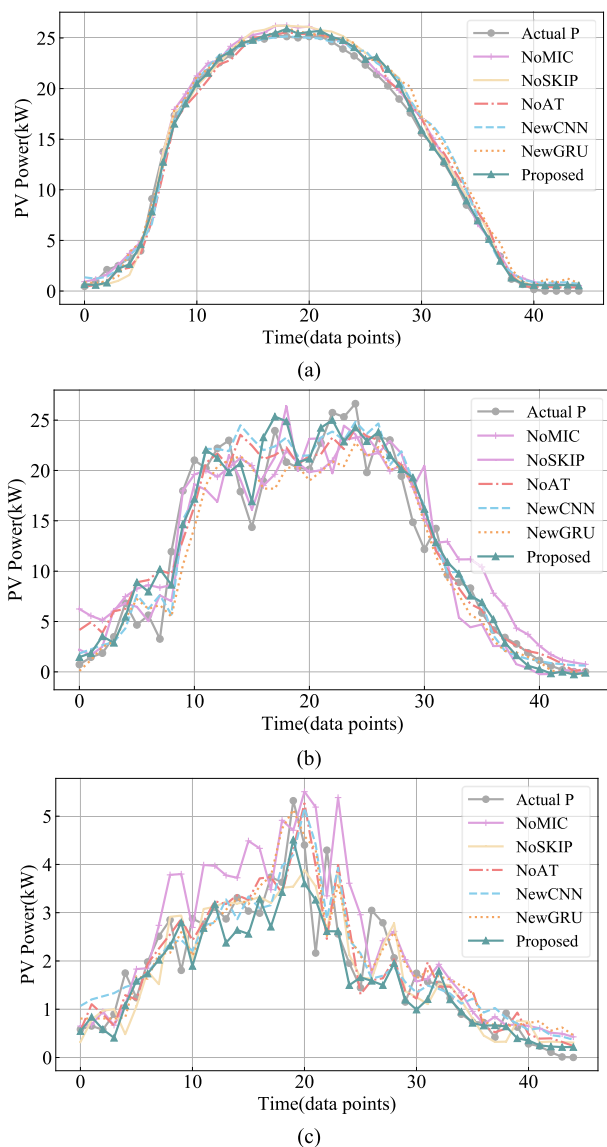


FIGURE 10. 15-minutes forecasting curves for three weather types under the ablation experiment. (a) Sunny. (b) Cloudy. (c) Rainy.

BiGRU and Attention, all kinds of forecasting indicators had stable improvements in all the weather, which verified the rationality of using BiGRU to learn deep bidirectional patterns and adopting Attention to optimize outputs.

To visually illustrate the forecasting outcomes, we present the power trend curves for a single test day across the three distinct weather conditions, as depicted in FIGURE 8. Notably, all the comparison models and the full model exhibited a strong fit to the one-day power trend in sunny conditions. However, the forecasting model that lacked the skip connection strategy displayed a slight deviation in the peak value. Under cloudy conditions, most of the sampling points could predict the power generation situations well. In particular, for the full model whose peaks and troughs were more consistent with the real values compared to the other, its

fluctuation trend basically matched the actual trend. In more complicated rainy weather, the full model could still reflect the overall pattern of power variation by using jump strategy and Attention, but some extreme points fluctuated too greatly to predict accurately. Meanwhile, similar samples were not selected by grey correlation degree, so the quality of the training samples was so low that the forecasting results were not ideal.

V. CONCLUSION

This paper proposes a phased PV power forecasting method suitable for a variety of complicated weather conditions, which can effectively take short-term nonlinear patterns and long periodicity in days of temporal data in a PV power process into account. In this method, similar samples were screened by MIC weighted gray correlation degree, parallel features and temporal patterns were extracted by the TCN-BiGRU model, longer-term information was captured by jump connection strategy, and, combining attention mechanism, periodic characteristics of sequences were fully utilized to further optimize the outputs. To verify the validity of the method, the comparison tests with mainstream models and ablation experiments of each module were carried out on real power station data sets. Consequently, they proved that the forecasting method was suitable for weather conditions with different fluctuations, and could effectively predict ultra-short-term power generation, with stable performance in predicting after 15 minutes, 30 minutes, 45 minutes, and 60 minutes. Thus, the experiment results indicate that the method can provide a data basis for photovoltaic grid connection and power station operation and maintenance. Since the data used in this paper can perform well on a time scale of 15 minutes to 60 minutes.

It should be pointed out that the meteorological data used in this paper for screening similar samples came from historical site data rather than historical forecast data. For the reason that historical meteorological forecast data are difficult to obtain as they normally would not be stored, and sometimes forecast data cannot accurately describe the meteorological state of a day. Therefore, when the method in this paper is applied to an actual forecasting project, the accuracy of the next day’s weather forecast can greatly affect the accuracy of forecasting. Nevertheless, the method in this paper does provide feasible forecasting means. Regarding the research in the next stage, we will focus on the following aspects:

1) Our approach employs the skip connection strategy that integrates short-term regularity and long-term periodicity. However, we do not conduct targeted training on linear regularity. In the realm of PV power generation forecasting, exploring methods to incorporate linear models into the network is a promising avenue for investigation. Doing so may enhance the resilience of the nonlinear deep learning model in the face of scale violations in time series data.

2) The model of this paper demonstrates excellent prediction performance when applied to electrical data with

a granularity of 15 minutes. We hold the belief that it may further excel when applied to even finer time intervals or analogous multivariate time series prediction scenarios. However, utilizing the method presented in this paper as the foundational model and fine-tuning or transferring learning by the specific situation remain formidable challenges.

3) As the precision of weather forecast data advances, the outcome of model SK-TCN-BILSTM can be employed as an initial PV forecast result. Theoretically, combining the forecast results obtained from our approach with the more precise weather forecast data could effectively handle photovoltaic energy generation in more intricate weather conditions. Nevertheless, the method for leveraging accurate forecast data to rectify the forecast results of our approach is a research direction that necessitates further investigation.

REFERENCES

- [1] A. Bermudez-Garcia, P. Voarino, and O. Raccurt, "Environments, needs and opportunities for future space photovoltaic power generation: A review," *Appl. Energy*, vol. 290, no. 5, May 2021, Art. no. 116757.
- [2] G. Etxegarai, A. López, N. Aginako, and F. Rodríguez, "An analysis of different deep learning neural networks for intra-hour solar irradiation forecasting to compute solar photovoltaic generators' energy production," *Energy Sustain. Develop.*, vol. 68, pp. 1–17, Jun. 2022.
- [3] C. Liu, M. Li, Y. Yu, Z. Wu, H. Gong, and F. Cheng, "A review of multitemporal and multispatial scales photovoltaic forecasting methods," *IEEE Access*, vol. 10, pp. 35073–35093, 2022, doi: [10.1109/ACCESS.2022.3162206](https://doi.org/10.1109/ACCESS.2022.3162206).
- [4] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, "Solar photovoltaic generation forecasting methods: A review," *Energy Convers. Manage.*, vol. 156, pp. 459–497, Jan. 2018.
- [5] Y. Ma, Q. Lv, R. Zhang, Y. Zhang, H. Zhu, and W. Yin, "Short-term photovoltaic power forecasting method based on irradiance correction and error forecasting," *Energy Rep.*, vol. 7, pp. 5495–5509, Nov. 2021.
- [6] X. Luo, D. Zhang, and X. Zhu, "Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge," *Energy*, vol. 225, Jun. 2021, Art. no. 120240.
- [7] M. J. Mayer and G. Gróf, "Extensive comparison of physical models for photovoltaic power forecasting," *Appl. Energy*, vol. 283, Feb. 2021, Art. no. 116239.
- [8] D. Z. Yang, E. Wu, and J. Kleissl, "Operational solar forecasting for the real-time market," *Int. J. Forecasting*, vol. 37, no. 3, p. 1325, Jun. 2020.
- [9] Z. Si, M. Yang, Y. Yu, and T. Ding, "Photovoltaic power forecast based on satellite images considering effects of solar position," *Appl. Energy*, vol. 302, Nov. 2021, Art. no. 117514.
- [10] Z. Zhao, J. Liu, Z. Zhang, F. Wang, H. Chai, Y. Yu, X. Lu, T. Wang, and Y. Lin, "Deep learning based surface irradiance mapping model for solar PV power forecasting using sky image," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 3385–3396, Aug. 2020.
- [11] H. X. Zang, L. Cheng, T. Ding, K. W. Cheung, Z. Liang, Z. Wei, and G. Sun, "Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network," *IET Gener., Transmiss. Distrib.*, vol. 12, no. 20, pp. 4557–4567, Nov. 2018.
- [12] P. Li, K. Zhou, X. Lu, and S. Yang, "A hybrid deep learning model for short-term PV power forecasting," *Appl. Energy*, vol. 259, Feb. 2020, Art. no. 114216.
- [13] Z. Qian, Y. Pei, H. Zareipour, and N. Chen, "A review and discussion of decomposition-based hybrid models for wind energy forecasting applications," *Appl. Energy*, vol. 235, pp. 939–953, Feb. 2019.
- [14] K. Cheng, L. M. Guo, Y. K. Wang, and M. T. Zafar, "Application of clustering analysis in the prediction of photovoltaic power generation based on neural network," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 93, Nov. 2017, Art. no. 012024.
- [15] B. Gu, H. Shen, X. Lei, H. Hu, and X. Liu, "Forecasting and uncertainty analysis of day-ahead photovoltaic power using a novel forecasting method," *Appl. Energy*, vol. 299, Oct. 2021, Art. no. 117291.
- [16] X. Ma and X. Zhang, "A short-term prediction model to forecast power of photovoltaic based on MFA-elman," *Energy Rep.*, vol. 8, pp. 495–507, Jul. 2022.
- [17] S. X. Wang, Y. M. Wang, Y. Liu, and N. Zhang, "Hourly solar radiation forecasting based on EMD and ELM neural network," *Electr. Power Autom. Equip.*, vol. 34, no. 8, pp. 7–12, Aug. 2014.
- [18] Q. Wang, S. Ji, M. Hu, W. Li, F. Liu, and L. Zhu, "Short-term photovoltaic power generation combination forecasting method based on similar day and cross entropy theory," *Int. J. Photoenergy*, vol. 2018, pp. 1–10, Dec. 2018.
- [19] B. Chen, P. Lin, Y. Lin, Y. Lai, S. Cheng, Z. Chen, and L. Wu, "Hour-ahead photovoltaic power forecast using a hybrid GRA-LSTM model based on multivariate meteorological factors and historical power datasets," *IOP Conf. Ser., Earth Environ. Sci.*, vol. 431, no. 1, Feb. 2020, Art. no. 012059.
- [20] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 300–308, Jan. 2018.
- [21] J. Boland, M. David, and P. Lauret, "Short term solar radiation forecasting: Island versus continental sites," *Energy*, vol. 113, pp. 186–192, Oct. 2016.
- [22] S. Ferrari, M. Lazzaroni, V. Piuri, L. Cristaldi, and M. Faifer, "Statistical models approach for solar radiation forecasting," in *Proc. IMTC*, 2013, pp. 1734–1739.
- [23] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting," *Appl. Sci.*, vol. 8, no. 1, p. 28, Dec. 2017.
- [24] M. K. Behera, I. Majumder, and N. Nayak, "Solar photovoltaic power forecasting using optimized modified extreme learning machine technique," *Eng. Sci. Technol., Int. J.*, vol. 21, no. 3, pp. 428–438, Jun. 2018.
- [25] Y. Zhou, N. Zhou, L. Gong, and M. Jiang, "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy*, vol. 204, Aug. 2020, Art. no. 117894.
- [26] M. Pan, C. Li, R. Gao, Y. Huang, H. You, T. Gu, and F. Qin, "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *J. Cleaner Prod.*, vol. 277, Dec. 2020, Art. no. 123948.
- [27] U. Munawar and Z. Wang, "A framework of using machine learning approaches for short-term solar power forecasting," *J. Electr. Eng. Technol.*, vol. 15, no. 2, pp. 561–569, Mar. 2020.
- [28] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, Feb. 2022, Art. no. 122812.
- [29] N. Lopac, F. Hrčić, I. P. Vuksanović, and J. Lerga, "Detection of non-stationary GW signals in high noise from cohen's class of time-frequency representations using deep learning," *IEEE Access*, vol. 10, pp. 2408–2428, 2022, doi: [10.1109/ACCESS.2021.3139850](https://doi.org/10.1109/ACCESS.2021.3139850).
- [30] A. A. Abdelhamid, E.-S.-M. El-Kenawy, B. Alotaibi, G. M. Amer, M. Y. Abdelkader, A. Ibrahim, and M. M. Eid, "Robust speech emotion recognition using CNN+LSTM based on stochastic fractal search optimization algorithm," *IEEE Access*, vol. 10, pp. 49265–49284, 2022, doi: [10.1109/ACCESS.2022.3172954](https://doi.org/10.1109/ACCESS.2022.3172954).
- [31] D. Kim, S.-W. Hwang, and J. Kim, "Very short-term photovoltaic power generation forecasting with convolutional neural networks," in *Proc. Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, Oct. 2018, pp. 1310–1312.
- [32] Y. Wang, W. Liao, and Y. Chang, "Gated recurrent unit network-based short-term photovoltaic forecasting," *Energies*, vol. 11, no. 8, p. 2163, Aug. 2018.
- [33] D. Lee and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, vol. 12, no. 2, p. 215, Jan. 2019.
- [34] H. Zhou, Y. Zhang, L. Yang, Q. Liu, K. Yan, and Y. Du, "Short-term photovoltaic power forecasting based on long short term memory neural network and attention mechanism," *IEEE Access*, vol. 7, pp. 78063–78074, 2019.
- [35] A. Mellit, A. M. Pavan, and V. Lughi, "Deep learning neural networks for short-term photovoltaic power forecasting," *Renew. Energy*, vol. 172, pp. 276–288, Jul. 2021.
- [36] A. Agga, A. Abbou, M. Labbadi, and Y. El Houm, "Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models," *Renew. Energy*, vol. 177, pp. 101–112, Nov. 2021.



HANGXIA ZHOU received the M.S. degree from Zhejiang University, Hangzhou, China. She is currently a Professor with the College of Information Engineering, China Jiliang University, Hangzhou. Her research interests include big analysis, machine learning, pattern recognition, and sustainable energy technologies. She is a Committee Member of the China Computer Federation (CCF).



CHEN CUI received the M.S. degree from China Jiliang University, Hangzhou, China, in 2015. He is currently pursuing the Ph.D. degree in computer science and technology with Hangzhou Dianzi University, Hangzhou. He is also currently a Lecturer in Zhejiang Police College. His current research interests include information hiding and image processing.



JUN WANG was born in Hangzhou, Zhejiang, China. He received the master's degree from China Jiliang University, Hangzhou. His research interests include data mining and time series forecasting.



FULIAN OUYANG was born in Yichun, Jiangxi, China. She received the master's degree from China Jiliang University, Hangzhou, China. Her research interests include smart grids and intelligent algorithm.



XIANBIN LI was born in Pujiang, Zhejiang, China. He is currently pursuing the master's degree in computer applications with China Jiliang University. His research interests include deep learning and natural language processing.

...