Applied Energy 391 (2025) 125869

Contents lists available at ScienceDirect

Applied Energy

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

A photovoltaic power forecasting framework based on Attention mechanism and parallel prediction architecture

Zhengda Zhou [a,b], Yeming Dai [a,*], Mingming Leng [c]

[a] School of Business, Qingdao University, Qingdao 200071, China
[b] School of Information Management, Nanjing University, Nanjing 210023, China
[c] Faculty of Business, Lingnan University, Hong Kong

HIGHLIGHTS

- A novel power forecasting framework with parallel architecture combining linear and nonlinear components is proposed.
- Introducing the Attention mechanism to assign weights to input variables according to their importance.
- A combined DCC-BiLSTM structure is designed to complementally capture both spatial and temporal features.
- An Autoregressive component is adopted to make prediction in parallel with nonlinear component.

ARTICLE INFO

Keywords:
Ultra-short-term photovoltaic power forecasting
Attention mechanism
Parallel prediction architecture
Autoregression
Dilated Causal Convolution-Bidirectional Long Short-Term Memory network

ABSTRACT

Photovoltaic power generation is susceptible to the stochastic volatility characteristics of meteorological conditions, so it is of great significance to forecast the photovoltaic power generation accurately and reliably. This paper proposes a novel hybrid forecasting framework (Attention-DCC-BiLSTM-AR model, ADBA model) for ultra-short-term photovoltaic power prediction, which combines the Attention mechanism and a well-designed parallel prediction architecture with linear Autoregressive (AR) component and nonlinear Dilated Causal Convolution-Bidirectional Long Short-Term Memory network (DCC-BiLSTM) component. Firstly, Attention mechanism is employed to assign weights to input variables according to their relative importance, so as to optimize the multivariate time series. Secondly, the optimized data is fed into linear and nonlinear components of the parallel architecture for prediction, respectively. The nonlinear prediction component is implemented by a combined DCC-BiLSTM structure, which has complementary strength in extracting spatial and temporal features. Subsequently, the extracted features are fed into feature mapping layers to obtain the nonlinear fitting results. The linear prediction component is implemented by a statistical AR model, which can mitigate the scale sensitivity problem associated with neural networks and provide the linear fitting results. This parallel prediction architecture enables the hybrid framework to model both linear and nonlinear characteristics of historical power generation time series simultaneously. Finally, the prediction results of two components are integrated to obtain the final prediction result. Experimental results demonstrate that: the proposed model consistently outperforms benchmark models in terms of forecasting accuracy and robustness, and has shown the most superior prediction performance on different sites, different seasons, and different prediction time horizons.

Abbreviations

NWP	Numerical Weather Prediction
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ANN	Artificial Neural Networks
RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory
PV	Photovoltaic
R_squared	R^2
SVM	Support Vector Machine

* Corresponding author.
E-mail addresses: yemingdai@163.com (Y. Dai), mmleng@ln.edu.hk (M. Leng).

https://doi.org/10.1016/j.apenergy.2025.125869
Received 14 August 2024; Received in revised form 20 February 2025; Accepted 3 April 2025
Available online 11 April 2025

DT	Decision Tree
ARMA	Autoregressive Moving Average
BiLSTM	Bidirectional Long Short-Term Memory
DCC	Dilated Causal Convolution
BiGRU	Bidirectional Gated Recurrent Unit
DKASC	Desert Knowledge Australia Solar Center
TCN	Temporal Convolutional Networks
RSE	Root Relative Squared Error
CORR	Empirical Correlation Coefficient
MAE	Mean Absolute Percentage Error
RMSE	Root Mean Square Error
RF	Random Forest
FC	Fully Connected Layer

1. Introduction

For the last few years, there has been a considerable upsurge in the global consumption of fossil fuels, which has led to severe environmental issues such as the worsening global warming and air pollution [1]. Furthermore, fossil fuels are non-renewable energy sources with limited reserves and are being depleted due to excessive human exploitation. In contrast, solar energy due to its superior characteristics like abundant reserves, low carbon emissions, and high flexibility, is gradually replacing traditional energy sources [2]. However, photovoltaic (PV) power systems are easily influenced by the stochastic volatility characteristics of meteorological conditions, resulting in fluctuations of generating power [3,4], which adversely affects the normal dispatching of grid system. Therefore, it is of great significance to forecast the PV power generation reliably and accurately.

Existing PV power prediction could broadly be categorized into three predominant types according to their prediction timescales. The first type is ultra-short-term prediction with its timescale ranging from a few seconds to a few minutes, which is suitable for the real-time security analysis and the real-time dispatching of grid system [5]. The second type is short-term prediction with its timescale ranging from one hour to one day, and is often be employed to balance electricity market transactions and adjust unit portfolio schemes [6]. The third type is medium or long-term prediction with its timescale ranging from day to week/month/year, which is primarily utilized for equipment maintenance and site selection of new power plants [7]. Besides, in terms of PV power prediction methods, they could be classified into four categories: physical methods, statistical methods, machine learning and deep learning methods, and hybrid methods [4].

For physical methods, the fundamental principles of hydrodynamics and thermodynamics are utilized to forecast meteorological changes. These methods try to study the relationship between PV power generation and meteorological variables, as well as the equipment installation configurations of PV power system. While physical methods like Numerical Weather Prediction (NWP) and image-based techniques [8,9] offer high interpretability and theoretical accuracy, making them suitable for data-scarce environments, but their complexity and computational demands may limit practicality in real-time or large-scale applications.

For statistical methods, based on principles of statistics, they extract effective information about the change patterns of power load from historical time series to make predictions. Statistical methods mainly include AR model [8,10], Autoregressive Moving Average (ARMA) [11], Autoregressive Integrated Moving Average (ARIMA) [12,13] and so on. Although statistical methods are simple and interpretable, their linear assumptions limit their ability to capture complex nonlinear relationships of multivariate input time series. Besides their prediction performance under unstable weather conditions (such as cloudy or rainy weather conditions) is not satisfactory [14].

With the advancements in artificial intelligence, machine learning and deep learning methods have surpassed traditional physical and statistical approaches in nonlinear fitting capabilities [15,16]. These methods excel in processing large-scale and complex data patterns, making them ideal for modern data-rich computational environments. Machine learning models primarily encompass Artificial Neural Networks (ANN) [17], Support Vector Machines (SVM) [18], Decision Trees (DT) [19] and so on. Deep learning models, such as Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN), offer more powerful capabilities in learning high-dimensional features and significantly reduce the need for feature engineering, enabling end-to-end nonlinear fitting and prediction. RNNs, with their memory capabilities and parameter-sharing mechanisms, are particularly effective for time series predictions [20]. Long Short-Term Memory (LSTM), as a variant of RNN, is widely used for extracting temporal features from historical data [5,21]. CNNs, meanwhile, excel at capturing spatial relationships between data points, making them suitable for multivariate inputs like historical power and meteorological variables [22,23].

Prediction methods based on above single model are often restricted by their limited learning ability and poor generalization ability, which makes achieving a more accurate and robust prediction performance become extremely challenging [24]. Alternatively, constructing a hybrid model combining the strengths of multiple forecasting techniques has been proven to be a feasible and effective method [25–27]. The hybrid models often integrate multiple different models to achieve the ensemble prediction outcomes with superior accuracy and robustness, such as hybrid prediction models combined LSTM and CNN [28,29], and hybrid prediction models combined CNN and BiGRU [30].

Although the previous hybrid PV prediction methods have already achieved relatively competitive prediction performance, they tend to overemphasize the application of neural networks of various architectures and only perform nonlinear fitting based on these neural networks without considering the linear characteristics of historical power generation time series [13]. Furthermore, existing methods overlook the issue that the scale of their output is insensitive to the input scale, that is, due to the inherent nonlinear nature of neural network, it prioritizes the periodic change characteristics at global scale [31], but relatively ignores the non-periodic change characteristics at the local scale [32]. In addition, in the scenarios of multivariate inputs, these methods tend to treat all features equally, without concerning the varying importance of different variables for PV prediction. These limitations impede further improvement of the model's predictive accuracy and stability.

Based on the aforementioned studies, this paper proposes a novel hybrid forecasting framework with superior accuracy and robustness, named Attention-DCC-BiLSTM-AR (ADBA) hybrid model, incorporating the Attention mechanism and a parallel prediction architecture with linear AR and nonlinear DCC-BiLSTM components for ultra-short-term PV power prediction. The prediction process can be divides into following steps: firstly, the Attention mechanism is employed to assign weights to the input variables according to their relative importance, so as to optimize the multivariate time series. Secondly, the optimized data are fed into linear component and nonlinear component of the parallel architecture for prediction, respectively. The nonlinear prediction component is implemented by neural network. In this component, a combined DCC-BiLSTM structure is designed to complementary extract spatial-temporal features from the input time series, and subsequently, the extracted features are fed into feature mapping layers to obtain the nonlinear fitting prediction results. The linear prediction component is implemented by a statistical model: AR model, which can mitigate the scale sensitivity problem associated with neural network and provide the linear fitting prediction results. Finally, the prediction results of the two components were integrated to obtain the final prediction result. To verify the superiority of the proposed model, real power generation data from different sites of DKASC (Desert Knowledge Australia Solar Center) is selected as the experimental data sources. Experimental results demonstrate that: the proposed ADBA model consistently outperforms benchmark models in terms of forecasting accuracy and robustness, and has shown the most superior prediction performance on different sites, different seasons, and different prediction time horizons. The main contributions of this paper are outlined in detail as follows:

(1) Introduce a novel approach incorporating an Attention mechanism for enhanced feature prioritization: We introduce the Attention mechanism into the ADBA model to dynamically assign weights to meteorological variables based on their importance. This allows the model to focus more on crucial features for PV power prediction, reducing the impact of less important variables and optimizing the input time series.
(2) Develop an advanced temporal and spatial feature extraction method using DCC-BiLSTM: We design a novel feature extraction component that combines the advantages of DCC in spatial feature extraction and strengths of BiLSTM in temporal dependency modeling. The combined structure shows distinct advantages in capturing both temporal and spatial features while preventing information leakage through its causal design.
(3) Propose a parallel linear and nonlinear modeling method for improved prediction accuracy: We propose a parallel prediction mechanism that integrates AR and DCC-BiLSTM components to enhance forecasting accuracy. This design leverages AR's proficiency in capturing linear patterns and DCC-BiLSTM's capability in modeling nonlinear dependencies, significantly improving the overall prediction performance.

The rest of this paper is organized as follows. Section 2 introduces the principles of different models in the proposed ADBA model in detail. Section 3 presents the components of the ADBA model and how they form the ADBA model. Section 4 provides the experimental process and experimental results of the case study. Section 5 draws conclusions and suggests directions for future work.

2. Methodology

This section goes into detail about the principles of different models in the ADBA hybrid forecasting framework.

2.1. Attention mechanism

In multivariate time series prediction, due to the various importance of different variables, the Attention mechanism can strengthen the model's attention to key variables and highlight the impact of significant information [33,34], while reducing the interference of redundant and low correlation variables, thereby optimizing the input time series and ameliorating prediction accuracy of the model. The structure of the Attention mechanism is shown in Fig. 1 [35].

In the Attention's structure shown in Fig. 1, $\{x_1, x_2, \ldots, x_t\}$denotes the input sequence, $\{h_1, h_2, \ldots, h_t\}$ denotes the hidden layer state, $\{a_1, a_2, \ldots, a_t\}$ denotes the attention weight of the hidden layer state, s_t denotes the product sum of the hidden state and the weight, and y_t is the output at time slot t. The specific calculation formula is as follows:

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

$$s_i = \sum_{i=1}^{t} a_i h_i \tag{2}$$

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

where e_i represents the information contained in the i^{th} hidden state, v_a^T and W_i represents corresponding weight.

2.2. Dilated causal convolution

Dilated Causal Convolution (DCC), based on causal convolutions, serves as the core component of Temporal Convolutional Networks (TCN) and was employed in capturing spatial dependencies in the proposed forecasting framework. Causal convolutions are specifically designed based on the principle of causality, where the state at time t in each layer is solely related to the state at time t and the state at time before t of last layer. Compared with ordinary convolution, causal convolution effectively eliminates the problem of information leakage, making it more suitable for time series prediction tasks [5].

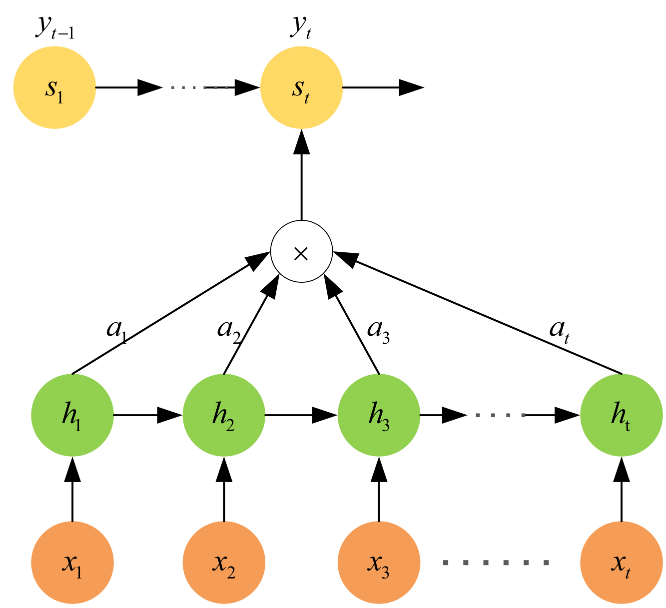

Fig. 1. Schematic diagram of the structure of the Attention mechanism.

Based on causal convolution, DCC increases the interval of convolution kernel by increasing the value of dilation coefficient, so that the network has a larger receptive field and allows longer historical time series as input [22,36], thereby capturing spatial features more effectively than CNN. As shown in Fig. 2, compared with causal convolution of the same number of layers, the receptive field of the DCC network expands from 4 to 8. The dilated convolution operation can be mathematically represented as:

$$\mathbf{F}(t) = \sum_{v=0}^{u-1} f(v) Xt - dv \tag{4}$$

where $\mathbf{F}(t)$ denotes a dilated convolution operation, $Xt - dv$ represents sequence data, $f(v)$ represents filter functions, u is the length of the input sequence data, v represents the number of the element in input sequence [34].

2.3. Long short-term memory network

The structure of the Long Short-Term Memory Network (LSTM) cell is depicted in Fig. 3. Based on RNN, LSTM incorporates internal memory unit and gate mechanism [37]. Compared with ordinary RNN, LSTM can accurately capture long-term temporal dependencies and alleviate the problem of gradient disappearance and gradient explosion [38,39]. The calculation of LSTM is as follows:

$$f_t = \sigma\left(W_f[a_{t-1}, x_t] + b_f\right) \tag{5}$$

$$f(t) = i_t = \sigma(W_i[a_{t-1}, x_t] + b_i) \tag{6}$$

$$\widetilde{C}_t = \tanh\left(W_g[a_{t-1}, x_t] + b_g\right) \tag{7}$$

$$C_t = f_t C_{t-1} + i_t \widetilde{C}_t \tag{8}$$

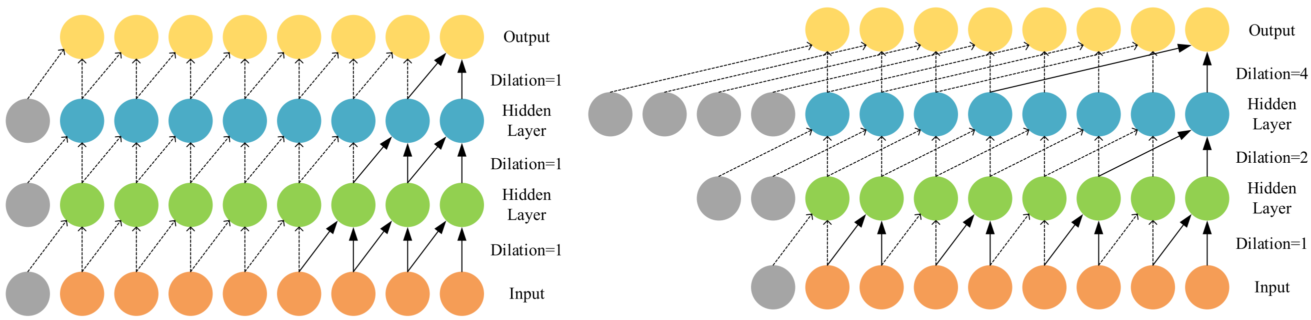

Fig. 2. Schematic diagram of causal convolution (left) and dilated causal convolution (right).

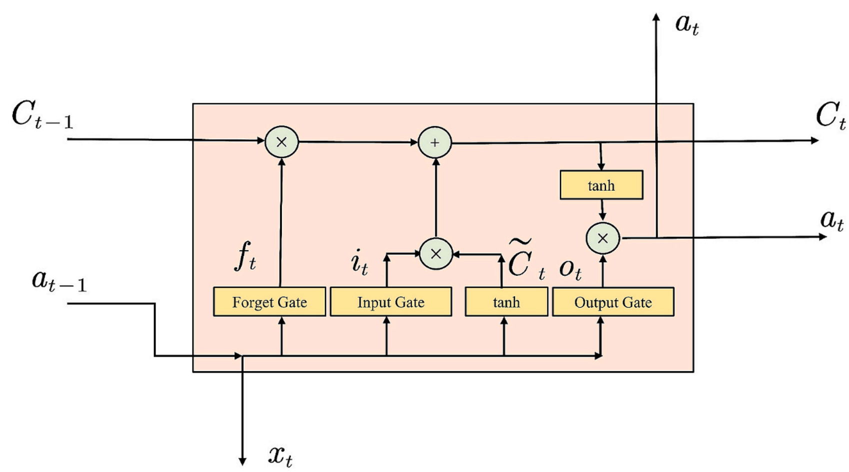

Fig. 3. Schematic diagram of the structure of LSTM cell.

$$o_t = \sigma(W_o[a_{t-1}, x_t] + b_o) \tag{9}$$

$$a_t = o_t \tanh C_t \tag{10}$$

where a_t represents the output of the hidden layer at time slot t, σ represents sigmoid activation function, W and b are respectively the weight matrix and bias term, C and tan$h$ represent the cell state and tangent activation functions, respectively [40,41].

As shown in Fig. 4, based on LSTM, BiLSTM consists of a forward LSTM layer and a backward LSTM layer [42], which can process the sequence forward and backward, respectively [43,44]. Compared with ordinary LSTM, BiLSTM can simultaneously consider the forward and backward information of the input time series to better capture the long-term temporal dependencies from historical power generation time series. The calculation of BiLSTM is as follows:

$$a_t^f = \mathrm{LSTM}\left(a_{t-1}^f, \mathbf{x}_t\right) \tag{11}$$

$$a_t^b = \mathrm{LSTM}\left(a_{t-1}^b, \mathbf{x}_t\right) \tag{12}$$

$$h_t^R = \mathbf{W}_{h^f} a_t^f + \mathbf{W}_{h^b} a_t^b + b \tag{13}$$

where $\mathrm{LSTM}(\cdot)$ denotes Eq. (14) to Eq. (19), a_t^f and a_t^b are the outputs of forward LSTM and backward LSTM at time slot t, respectively. $\mathbf{W}_{h^f}$

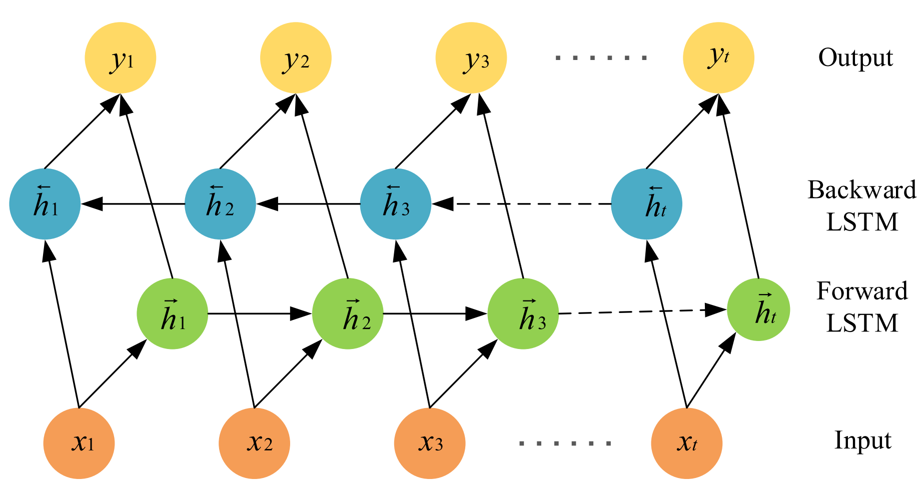

Fig. 4. Schematic diagram of the structure of BiLSTM.

and $\mathbf{W}_{h^b}$ represent the weights of forward and backward hidden sequences, respectively. b is the bias term; h_t^R is the final output of BiLSTM at time slot t.

2.4. Feature mapping

The feature mapping component is used for mapping the features extracted from the original data to obtain the nonlinear fitting prediction results of the neural network. It is generally implemented by the fullly connection layers (FC). The structure of feature mapping component is illustrated as Fig. 5, and its calculation process is as follows:

$$h_t^D = W^R h_t^R + b \tag{14}$$

where h_t^D is the prediction result of the neural network component at time t, and b is the bias term of the output layer.

2.5. Autoregressive (AR) model

The Autoregressive (AR) module is a classical statistical model frequently applied in time series prediction tasks across various disciplines, such as statistics and econometrics. It is particularly effective for capturing non-periodic change characteristics by focusing on local variations. The AR model assumes that prediction outcomes depend linearly on recent historical data, making it well-suited for modeling changes at the local scale. This linear fitting approach allows the AR module to effectively identify and predict non-periodic fluctuations in natural photovoltaic power time series which are often characterized by short-term, irregular changes [45,46].

The calculation formula of the Autoregressive model is as follows:

$$h_t^L = \sum_{k=0}^{q^{ar}-1} W_k^{ar} y_{t-k} + b^{ar} \tag{15}$$

where h_t^L denotes the prediction result of the Autoregressive component at time slot t, W^{ar} and b^{ar} are the weight coefficient and bias term of the AR model, respectively, and q^{ar} is the size of the input window of the AR component [32].

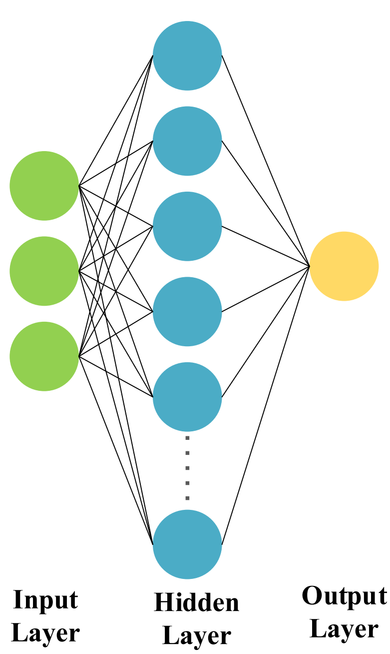

Fig. 5. Schematic diagram of the feature mapping component.

3. ADBA ultra-short-term PV power prediction model

Based on the fundamental models mentioned in Section 2, this section will provide an in-depth presentation of ADBA ultra-short-term PV power prediction model's the components and how they form the ADBA model.

3.1. Attention mechanism for feature weighting

Instead of treating all variables equally, the ADBA model proposed in this paper incorporates the Attention mechanism in order to assign weights to different meteorological features according to their relative importance. This allows the model to prioritize the more important features and alleviate the interference of redundant and low correlation features, thereby optimizing the input time series and setting a solid foundation for subsequent prediction process. At the same time, due to the large scale of original multivariate time series, in order to improve the training efficiency and reduce computational costs, the classical Attention mechanism is simplified by adopting the idea of multi-class logistic regression to implement the weight assignment of input information [33]. The principle of multi-class logistic regression is depicted in Fig. 6, where c_{nn} represents the connection weight between two neurons.

The overall output of this component is

$$\boldsymbol{\alpha} = f(\mathbf{C}\boldsymbol{X} + \mathbf{b}) \tag{16}$$

where $\boldsymbol{\alpha}$ denotes the attention allocation weight vector, $\mathbf{C}$ denotes the connection weight matrix, and $\mathbf{b}$ denotes the bias vector. f denotes the softmax function, and the weight allocated to the n^{th} input vector is calculated as follows:

$$\alpha_n = \frac{\exp(\mathbf{c}_n^{\mathrm{T}}\mathbf{X} + \mathbf{b}_n)}{\sum_{j=1}^{N} \exp\left(\mathbf{c}_j^{\mathrm{T}}\mathbf{X} + \mathbf{b}_j\right)} \tag{17}$$

where $\boldsymbol{c}_n$ is the connection weight vector between the n^{th} input vector and the rest vectors, and $\boldsymbol{b}_n$ is the bias term of the n^{th} input vector.

3.2. Combined DCC-BiLSTM nonlinear prediction component

In this work, a combined DCC-BiLSTM structure is designed for feature extraction in the neural network module. DCC has shown excellent performance in extracting spatial features. Compared with ordinary CNN, DCC can prevent the problem of temporal information leakage and allow longer historical time series as input which makes the network obtain a larger receptive field. Compared with ordinary LSTM, BiLSTM can simultaneously consider the forward and backward information of the input time series so that it has better capability in capturing long-term temporal dependency, and also can be employed to

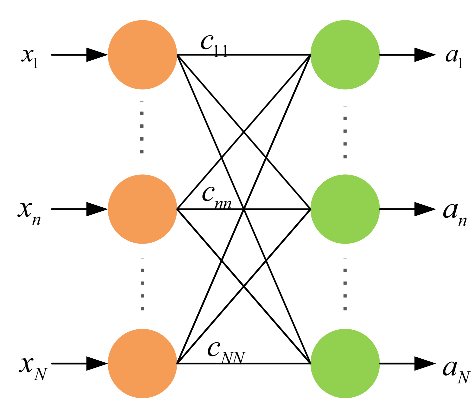

Fig. 6. Schematic diagram of the multi-class logistic regression.

compensate for the shortcomings of DCC, which emphasizes the extraction of spatial information while ignoring temporal information. The combined DCC-BiLSTM feature extraction structure fully combines the respective advantages of DCC subcomponent and BiLSTM subcomponent, and has exceptional ability in capturing both temporal and spatial features of time series. Subsequently, the extracted features are fed into the mapping layers to obtain the nonlinear fitting prediction results of the neural network. The detailed configurations of DCC subcomponent and BiLSTM subcomponent of combined DCC-BiLSTM nonlinear prediction component in the proposed framework are introduced as follows:

3.2.1. DCC subcomponent

DCC subcomponent consists of a three DCC blocks. Each layer of the DCC block is depicted in Fig. 7. The dilation coefficients of the three blocks are set to 1, 2, and 2, respectively. The specific parameter settings for the DCC subcomponent can be found in Table 1.

3.2.2. BiLSTM subcomponent

BiLSTM subcomponent utilizes a two-layer BiLSTM architecture and the configurations of the two layers are identical (refer to Fig. 4 in Section 2 for the structure of single-layer BiLSTM). The specific parameter settings for BiLSTM subcomponent can be found in Table 2.

3.3. AR linear prediction component

Since DCC and BiLSTM are both neural networks which primarily focus on nonlinear fitting, they may overlook linear and non-periodic pattern in the historical power generation time series, resulting in the predicted output scale being insensitive to input scale [32,47].To address this issue, this paper innovatively designs a linear fitting component which focuses on the change characteristics at the local scale, to make prediction in parallel with the DCC-BiLSTM nonlinear prediction component. In ADBA model, a statistical model called Autoregressive model is employed to implement this linear prediction component. The implementation mechanism of the AR model is described in detail in Section 2.5. For more information about it, please refer to that section.

The final prediction outcome of the ADBA model is obtained by the point-wise addition [48–50]49,50 of the prediction results from the DCC-BiLSTM nonlinear prediction component and the AR linear prediction component. Formally, this can be expressed as

$$\widehat{y}_t = h_t^D + h_t^L \tag{18}$$

where $\widehat{y}_t$ denotes the final prediction result of the ADBA model at time slot t. h_t^D and h_t^L represent the prediction result of DCC-BiLSTM component and AR component, respectively. This parallel prediction architecture combines the advantages of both types of components and enables the ADBA framework to model both linear and nonlinear characteristics of historical power generation time series simultaneously.

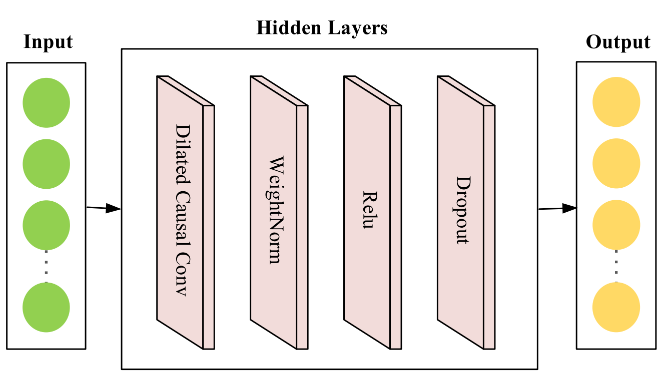

Fig. 7. The structure of single DCC block.

Table 1
Parameter settings of DCC subcomponent.

Layer	kernel_size	dilation	padding	dropout rate
DCC block 1	2	1	1	0.2
DCC block 2	2	2	2	0.2
DCC block 3	2	2	2	0

Table 2
Parameter settings of BiLSTM subcomponent.

Layer	input_size	hidden_size	batch_first	dropout
BiLSTM layer 1	128	50	True	0
BiLSTM layer 2	128	50	True	0

3.4. The overall framework of ADBA hybrid prediction model

Based on the aforementioned contents, the Attention-DCC-BiLSTM-AR hybrid power prediction model (ADBA model for short) is constructed, and the overall framework of the model is depicted as Fig. 8. Firstly, the Attention mechanism is employed to assign weights to the input variables according to their different importance, thereby optimizing of the multivariate time series. Secondly, the optimized data are fed into linear component and nonlinear component of the parallel architecture for prediction, respectively. The nonlinear prediction component is implemented by a combined DCC-BiLSTM structure, which has complementary strength in extracting spatial and temporal features. Subsequently, the extracted features are fed into feature mapping layers to obtain the nonlinear fitting results. The linear component is implemented by a statistical model: Autoregressive model, which can mitigate the scale sensitivity problem associated with neural network and provide the linear fitting prediction results. This parallel prediction architecture enables the ADBA framework to model both linear and nonlinear characteristics of historical power generation time series simultaneously. Finally, the prediction results of the two components are aggregated to obtain the final prediction result.

4. Case study

In this section, a series of horizontal contrast experiments are conducted to validate the superiority of the proposed PV power prediction model. The content of this section includes the ultra-short-term PV power prediction process based on ADBA model, the introduction of experimental equipment and model parameters, the description of data sources and preprocessing methods, feature selection, benchmark models, and analysis & discussion of experimental results.

4.1. Ultra-short-term photovoltaic power prediction process based on ADBA model

The overall process of ultra-short-term PV power prediction based on the ADBA model is shown as Fig. 9, which mainly involves the following steps:

Step 1: Data preparation. Firstly, the original multivariate time series data from photovoltaic power sites undergoes data preprocessing, which includes filling missing values and data normalization. Secondly, the original features are selected based on Pearson correlation coefficient, Kendall correlation coefficient and Spearman correlation coefficient. To reduce redundancy and achieve dimensionality reduction, features with low correlation are eliminated. Thirdly, the data after feature selection is processed by a sliding window algorithm to be transformed into an acceptable input format for ADBA model.

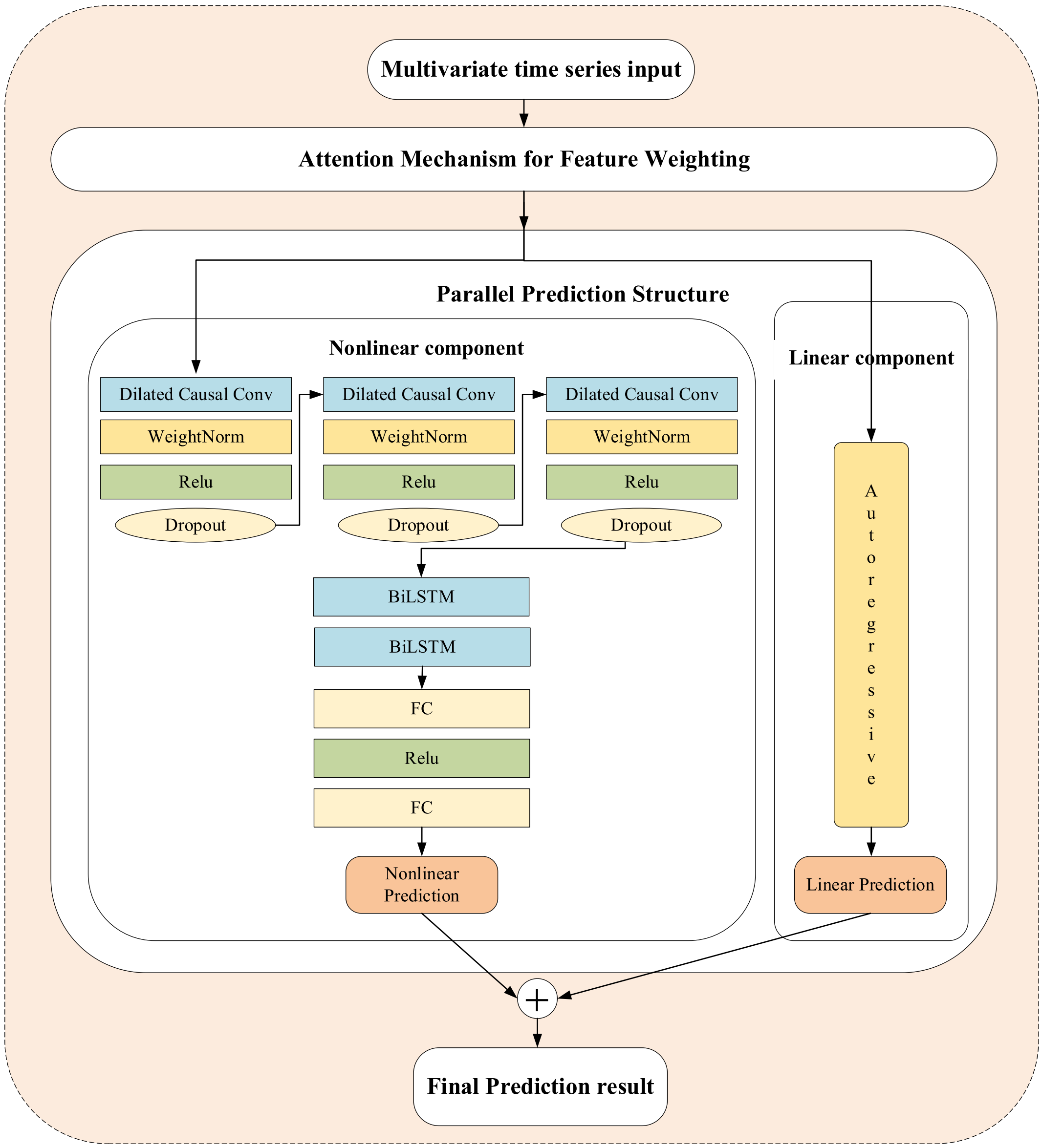

Fig. 8. Overall framework of the ADBA hybrid prediction model.

Step 2: Model Prediction. The processed multivariate time series is fed to the ADBA model to obtain the final prediction result. Refer to Section 3.4 for a detailed model prediction process.

Step 3: Performance evaluation. Five metrics are employed to measure the model prediction performance. At the same time, in order to validate the superiority of the proposed model, this paper also conducts a series of horizontal contrast experiments, that is, by comparing the prediction performance of ADBA model with other five benchmark models on different experimental sites, different seasons and different prediction horizons to corroborate that the proposed ADBA model has better prediction accuracy and prediction robustness than other models.

4.2. Experimental equipment and model parameters

All experiments in this paper were conducted on a PC with AMD Ryzen 7 4800H processor (with Radeon Graphics 2.90GHz) and 8GB RAM. The experimental models' architectures were implemented using Python 3.8.6 and Pytorch 2.0.0. The DCC-BiLSTM and AR modules were trained simultaneously within an integrated framework. The mean absolute error (MAE) loss between the predicted and actual results served as the supervisory signal for parameter updates, and the Adam optimizer was chosen as the model optimization algorithm. Additionally, the hyperparameters of the models were determined using a grid search algorithm. The optimal configuration was selected based on performance metrics evaluated on a validation set, ensuring robust and effective model performance.

4.3. Datasets description

To verify the effectiveness and robustness of the model proposed in this paper, and to ensure the representativeness and dependability of data source, the real power generation data and meteorological data from Desert Gardens (Site 1) and Service Station (Site 2) with different meteorological characteristics of the Yulara Solar System in Desert Knowledge Australia Solar Centre (DKASC) were selected as data sources to train and test model. DKASC is a widely used source of photovoltaic power data for research purposes. It is composed of multiple power sites with various meteorological conditions and data patterns. The geographical location of the experimental sites and the detailed description of their equipment installation configuration are shown in Fig. 10 and Table 3, respectively. The original features include Active Power, Wind Speed, Weather Temperature Celsius and so on (as shown

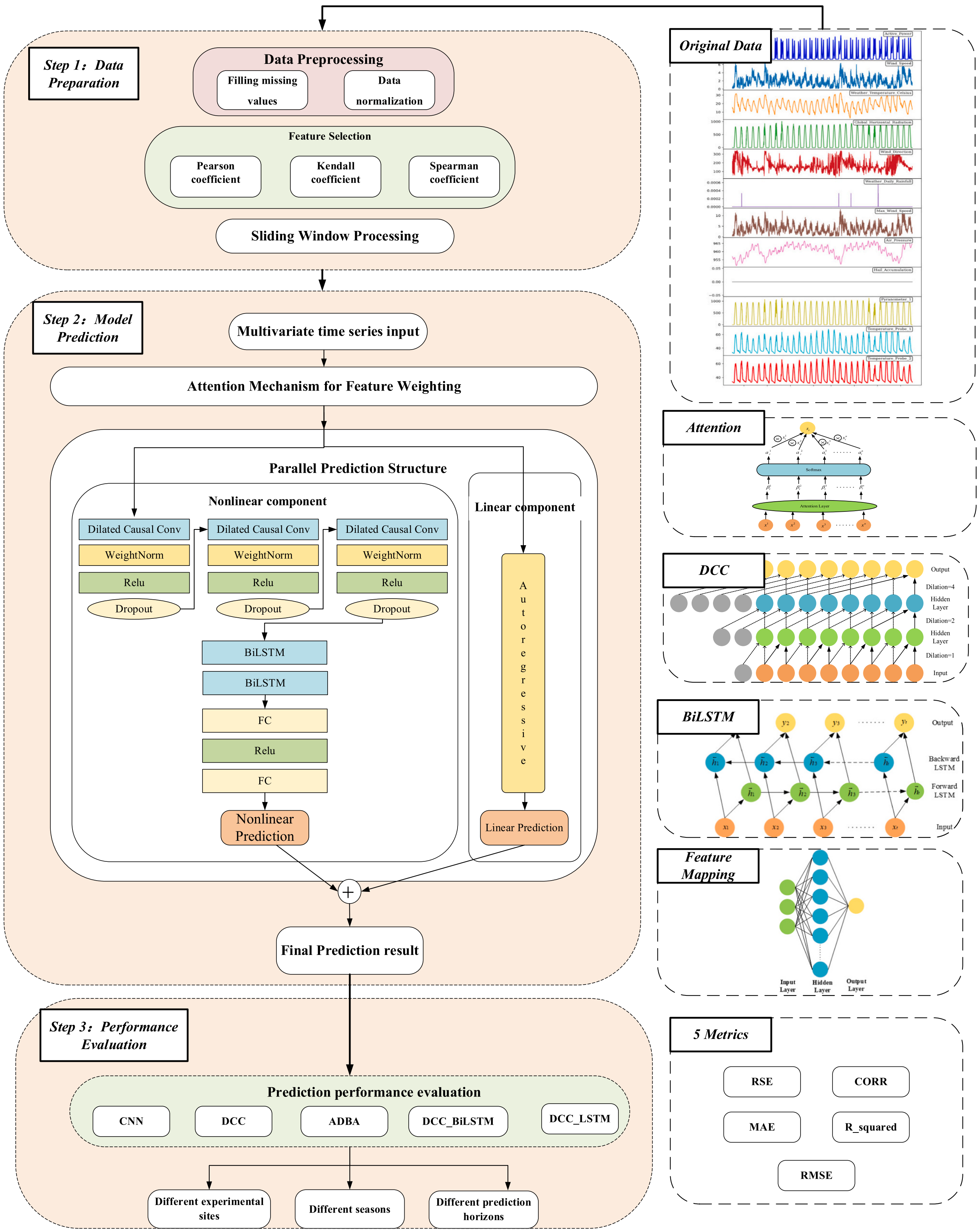

Fig. 9. The overall process of the ultra-short-term PV power prediction based on the ADBA model.

in Fig. 11). The data were collected at a sampling interval of 5 minutes.

4.4. Data preprocessing

Because raw data from real photovoltaic power stations is often crude and unsuitable as the direct input to the model, data preprocessing

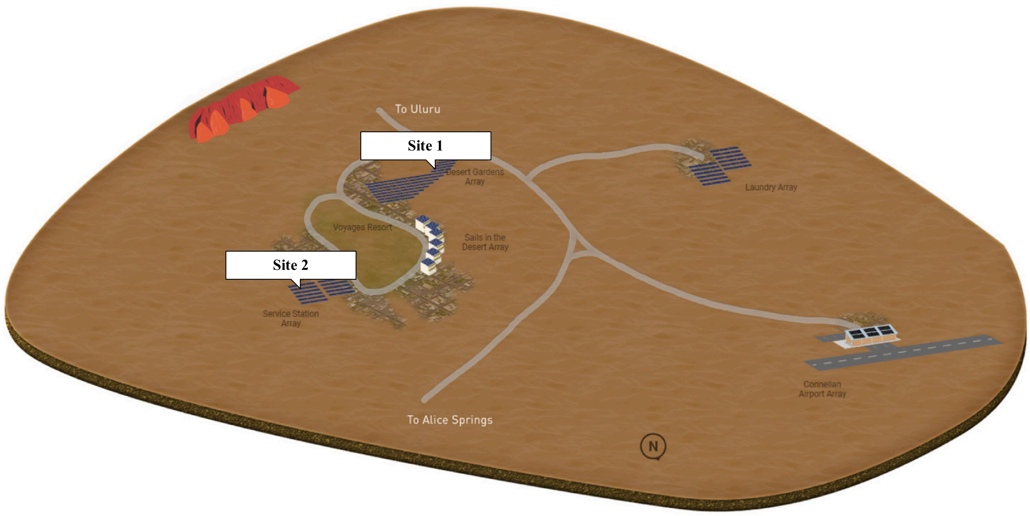

Fig. 10. The geographical location map of the experimental sites.

Table 3
Detailed equipment installation configuration of the experimental sites.

	Site.1 Desert Gardens	Site.2 Service Station
Array Rating	1058.4 kW	226.8 kW
PV Technology	poly-Si	poly-Si
Array Structure	Fixed: Ground Mount	Fixed: Ground Mount
Installed	2016	2016
Panel Rating	315 W	315 W
Number Of Panels	3360	720
Panel Type	Jinko Solar JKM315PP	Jinko Solar JKM315PP
Inverter Size / Type	SMA STC 25000TL-30	SMA STP 25000TL-30
Installation Completed	Thu, 3 Mar 2016	Thu, 3 Mar 2016
Array Tilt/Azimuth	Tilt = 15, Azi = 0 (Solar North)	Tilt = 15, Azi = 0 (Solar North)

has become an indispensable part of the data mining process in order to provide refined data for the prediction model [23]. Data preprocessing in this paper mainly includes two steps: filling missing value and data normalization. Firstly, the interpolation method was employed to fill in the missing values, so as to ensure the continuity of the time series data. Secondly, after the missing values are filled in, since the value range of different features vary significantly which is not conducive to model training, in order to accelerate the model's convergence and unify the scales of features, "Min-Max" Normalization method is employed to normalize the data. The normalization formula is shown in Eq. (33).

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{19}$$

where x denotes the data before normalization and x' denotes the data after normalization.

Besides, due to the special requirements of the time series prediction tasks, the original time series data must be processed by sliding window to become usable training samples. Therefore, a sliding window processing is conducted to convert the data to an "input-output" pairs format. This format allows data to be properly utilized as the input of the ADBA model during the training process.

4.5. Feature selection

After removing irrelevant variables, the original data consists of 11 characteristic variables and one target variable (Active Power). To further refine the dataset and eliminate redundant information and noise, Pearson correlation coefficient, Kendall correlation coefficient and Spearman correlation coefficient were applied to investigate the correlation between 11 characteristic variables and the target variable, and the heat map was drawn for visualization. As shown in Table 4 and Fig. 12, the values of three correlation coefficients between the variable Hail_Accumulation and the target variable Active_Power are all close to 0, indicating that Hail_Accumulation has a relatively low predictive importance. Therefore, this variable is removed for the purpose of dimensionality reduction and the remaining 10 characteristic variables are retained for model prediction.

4.6. Benchmark models

In order to verify the superiority of the proposed model, Convolutional Neural Network (CNN), Dilated Causal Convolution (DCC), AR (Autoregressive model), Dilated Causal Convolution-Long Short-Term Memory Network (DCC-LSTM), Dilated Causal Convolution-Bidirectional Long Short-Term Memory Network (DCC-BilSTM) are selected as five benchmark models to make horizontal comparison with the ADBA model proposed in this paper. And it is guaranteed that the experimental parameter settings of the five benchmark models are identical to those of the ADBA model.

4.7. Performance metrics

In order to quantify the prediction performance of models, five common evaluation metrics are employed. The five metrics are Root Relative Squared Error (RSE), Empirical Correlation Coefficient (CORR), Mean Absolute Percentage Error (MAE), R^2 (R_squared), Root Mean Square Error (RMSE). Among them, RSE, MAE, and RMSE measure the discrepancy between the actual and predicted values, and thus lower values are desirable. On the other hand, CORR and R_squared assess the closeness between the actual and predicted values, and higher values indicates better performance. The formulas of five metrics are as follows [51]:

$$RSE = \frac{\sqrt{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^{n} (y_i - \overline{y})^2}} \tag{20}$$

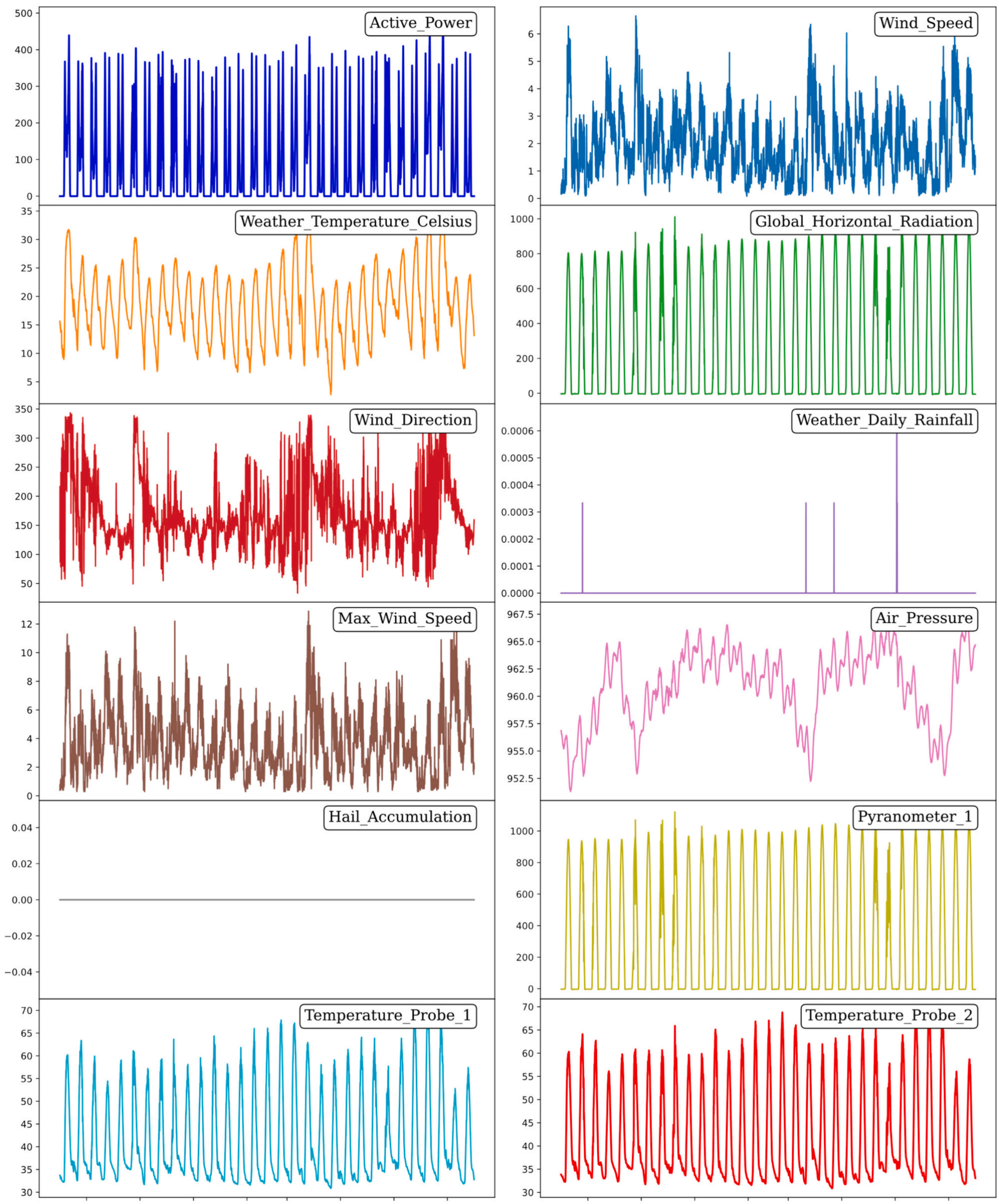

Fig. 11. The 12 original features.

$$\mathrm{CORR} = \frac{\frac{1}{n}\sum_{i=1}^{n}(y_i - mean(y_i))(\widehat{y}_i - mean(\widehat{y}))}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - mean(y_i))^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{y}_i - mean(\widehat{y}\right)^2}} \tag{21}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| \tag{22}$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} \tag{23}$$

Table 4
Results of correlation analysis between 11 feature variables and target variable (Active Power).

Variable	Pearson correlation coefficient	Kendall correlation coefficient	Spearman correlation coefficient
Temperature_Probe_2	0.703462	0.481386	0.671787
Temperature_Probe_1	0.700321	0.483437	0.673634
Global_Horizontal_Radiation	0.534287	0.597224	0.801963
Pyranometer_1	0.023486	0.585931	0.788238
Hail_Accumulation	**−0.00047**	**−0.00078**	**−0.00096**
Weather_Daily_Rainfall	−0.00049	−0.02529	−0.03103
Max_Wind_Speed	−0.00054	0.354863	0.519658
Weather_Temperature_Celsius	−0.00058	0.288498	0.403639
Wind_Speed	−0.00058	0.312829	0.463964
Air_Pressure	−0.00058	−0.07089	−0.10294
Wind_Direction	−0.00058	−0.11746	−0.17224

absolute prediction error of DCC is relatively reduced, which confirms that the larger receptive field of DCC can improve the prediction performance. Compared with CNN and DCC, both DCC-LSTM and DCC-BiLSTM achieve relatively high prediction accuracy, which reflects the advantages of LSTM and BiLSTM in extracting time series information and they can compensate for the shortcomings of CNN and DCC (which emphasize the extraction of spatial characteristics while ignoring temporal characteristics). This also illustrates the effectiveness of the combined DCC-BiLSTM feature extraction structure. Notably, the AR model demonstrates superior performance, achieving the most accurate prediction on the third day, which indicates the significance of linear components in the prediction task. Compared with DCC-BiLSTM, the ADBA model further improved the prediction accuracy. The MAE of ADBA model decreased by 19 % on the first day, 17 % on the second day and 4 % on the third day, which demonstrates that the Attention mechanism for feature weighting and the AR linear prediction component can improve the prediction performance of the model. On the entire test set, as shown in Table 6, all metric evaluation results of ADBA model are better than other benchmark models.

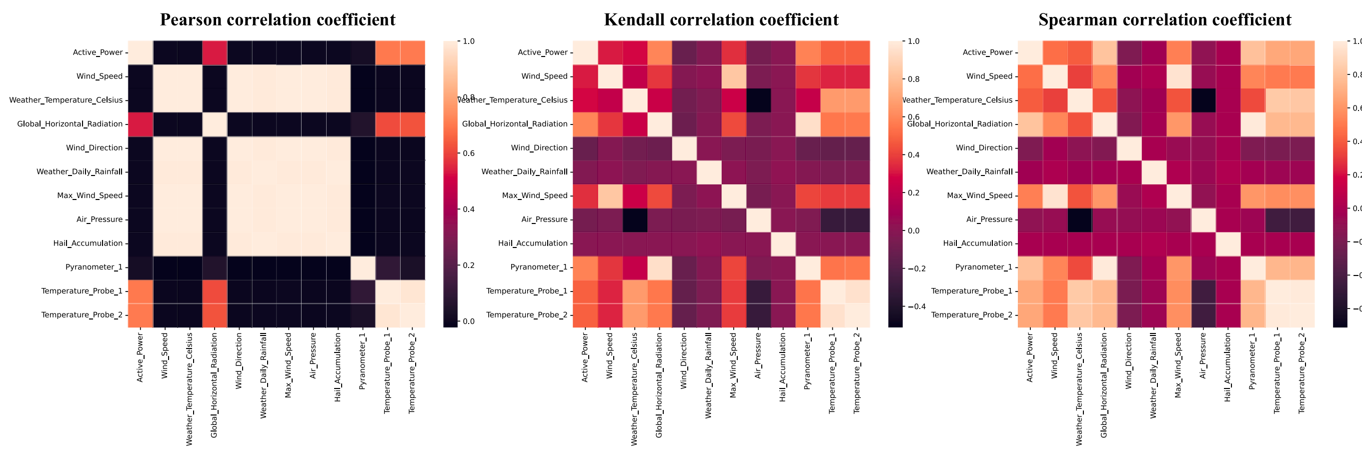

Fig. 12. Heat map of correlation among variables.

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

where y_i denotes the actual value, $\hat{y}_i$ denotes the predicted value, and n is the number of samples in test set.

4.8. Horizontal comparison of model prediction performance on different sites

In order to compare the prediction performance of the proposed ADBA model with other five benchmark models, data from PV Site 1 with a more complex meteorological pattern (also shows obvious linear characteristics) and data from PV Site 2 with a simpler meteorological pattern (also shows obvious nonlinear characteristics) were selected for experimental data sources. The data sampling time range was one month from August 15, 2017 to September 15, 2017. Following [14], data from the first 20 days was used as the training set, data from the subsequent five days was used as the validation set, and data form the remaining days was used as the test set. The models will predict PV power at future time points based on the data from the past 24 consecutive time points. Finally, the prediction performance of the five models was evaluated according to their performance on the test set.

For data from Site 1 with a more complex meteorological pattern and obvious linear characteristics, a three-day subset was selected from the test set to observe the prediction performance of the five models. As shown in Fig. 13, Fig. 14 and Table 5, the ADBA model shows the best prediction performance on all days except the third day, where it is slightly inferior to the AR model. CNN model exhibits the worst prediction performance, except on the first day. Compared with CNN, the

To further assess the generalization capability and robustness of the proposed model, data from Site 2 was selected as the second experimental data source due to its different meteorological pattern and stronger nonlinear characteristics (we employed the Hurst exponent (H)[1] to quantify the linear/nonlinear characteristics of different time series. Our analysis shows that Site 1 (H = 0.268) exhibits relatively more linear patterns, while Site 2 (H = 0.146) demonstrates stronger nonlinear characteristics.). Similarly, a three-day subset from the test set was selected to specifically observe the prediction performance of the models. As shown in Fig. 15, Fig. 16 and Table 7, the ADBA model achieves the best prediction performance on all three days, and also achieves the best overall performance on the entire test set as shown in Table 8.

The performance comparison between different models also provides insights into their respective strengths. On Site 1, with its more linear characteristics, the AR model achieves comparable results (RMSE = 24.69, R^2 = 0.958) to DCC-BiLSTM (RMSE = 25.56, R^2 = 0.955). In contrast, on Site 2 where nonlinear patterns dominate, DCC-BiLSTM shows better performance (RMSE = 13.64, R^2 = 0.966) than AR (RMSE = 15.01, R^2 = 0.958). These results align with the theoretical expectations of each component's capabilities.

[1] The Hurst exponent [Hurst H E. Long-term storage capacity of reservoirs[J]. Transactions of the American Society of Civil Engineers, 1951, 116: 770–799] measures the long-term memory of time series, ranging from 0 to 1, where H < 0.5 indicates stronger nonlinear characteristics.

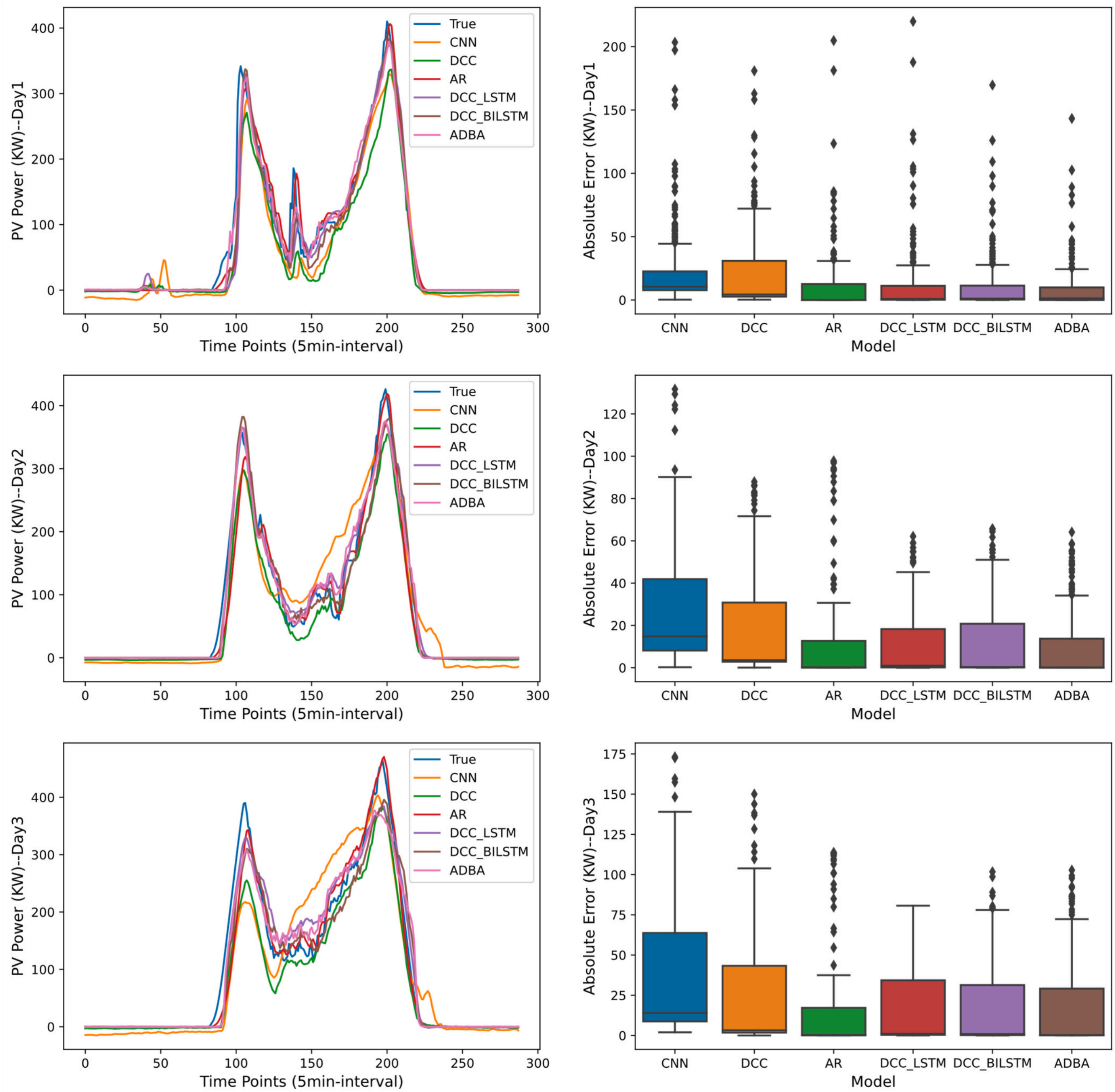

Fig. 13. Prediction performance of ADBA model and five benchmark models on Site 1 for three days.

Among other benchmark models, CNN shows considerable prediction instability, performing well on the first day but deteriorating significantly on the following days. DCC and DCC-LSTM maintain relatively stable prediction accuracy throughout the test period. The ADBA model, leveraging both linear and nonlinear modeling capabilities, achieves the best overall performance across both sites. Specifically, compared to DCC-BiLSTM, the ADBA model reduces MAE by 6 %, 35 %, and 39 % on the first, second, and third days respectively.

4.9. *Horizontal comparison of model prediction performance on different seasons*

To further verify the stability and robustness of the model's prediction performance and evaluate its applicability under different seasonal weather conditions, the original data of Site 1 was divided into four separate datasets: spring, summer, autumn and winter. One month of data was selected from each dataset for the experiment. Similarly, the first twenty days data of each dataset was applied for model training, the following five-days data was applied for model validation, and the remaining day data was applied for model testing. The overall prediction performance of the ADBA model and five benchmark models on four seasons is shown in Table 9-12 and Fig. 17. The ADBA model outperforms other benchmark models across almost all seasons, except on the summer data set, where it is slightly inferior to the DCC (DCC exhibits poorer prediction performance compared with the ADBA model in the remaining three seasonal datasets. Additionally, the prediction performance of the DCC model varies significantly across different seasons, indicating its noticeable instability). The experimental results on four seasonal datasets indicate that the ADBA model has better prediction stability and generalization ability.

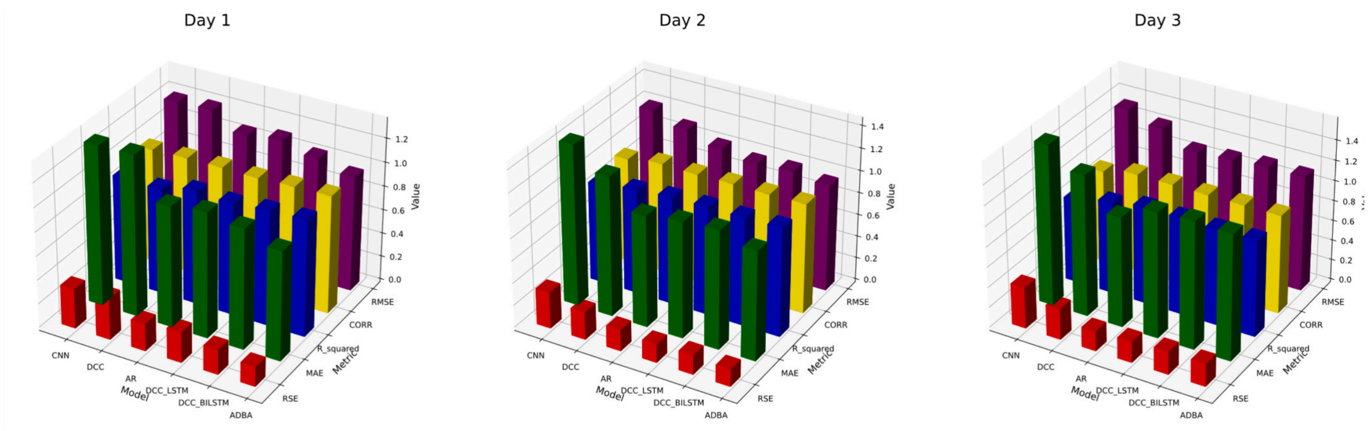

Fig. 14. The prediction performance metrics of ADBA model and five benchmark models on Site 1 for three days. (Note: We applied log10 transformation for MAE and log20 transformation for RMSE in advance, so that the scales of different metrics can match)

Table 5
The prediction performance metrics of ADBA model and five benchmark models on Site 1 for three days.

	Metrics	CNN	DCC	AR	DCC_LSTM	DCC_BiLSTM	ADBA
Day1	RSE	0.3328	0.3434	0.2293	0.256	0.2074	**0.1655**
	CORR	0.9605	0.9697	0.9735	0.9679	0.9794	**0.9867**
	MAE	21.571	21.9538	10.1860	11.1206	10.0342	**8.116**
	R_squared	0.8893	0.8821	0.9474	0.9345	0.957	**0.9726**
	RMSE	35.851	36.9911	24.7014	27.5775	22.3434	**17.8341**
Day2	RSE	0.328	0.2485	0.1913	0.1659	0.1746	**0.1548**
	CORR	0.9457	**0.9889**	0.9822	0.9865	0.9847	0.9881
	MAE	27.569	17.8304	10.0055	10.9437	11.4904	**9.4979**
	R_squared	0.8924	0.9383	0.9634	0.9725	0.9695	**0.976**
	RMSE	37.4274	28.3569	21.8290	18.9288	19.9245	**17.6652**
Day3	RSE	0.4089	0.3065	**0.1959**	0.2081	0.2267	0.2244
	CORR	0.9167	**0.9829**	0.9810	0.9782	0.9768	0.9751
	MAE	38.4757	24.5757	**12.1487**	17.1907	17.9157	17.216
	R_squared	0.8328	0.9061	0.9016	0.9567	0.9486	**0.9496**
	RMSE	55.0804	41.2898	**26.3879**	28.0409	30.5372	30.236

Table 6
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1.

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.4212	0.9077	34.7414	0.8226	50.8
DCC	0.3465	0.9563	25.4325	0.88	41.7877
AR	0.2047	0.9796	12.6478	0.9581	24.6939
DCC_LSTM	0.2227	0.9751	15.3235	0.9504	26.8625
DCC_BiLSTM	0.2119	0.9781	14.5499	0.9551	25.5602
ADBA	**0.1891**	**0.9825**	**12.1837**	**0.9642**	**22.8057**

4.10. Horizontal comparison of model prediction performance on different time horizons

In order to compare the performance of prediction models on different time horizons, the original data from Site 1 which had a sampling interval of 5 minutes, was resampled into other two different sampling interval types: 10 minutes and 15 minutes. Additionally, to ensure that the model have enough training samples, the time range of data sampling was extended from one month from August 15, 2017 to September 15, 2017 to two months from July 15, 2017 to September 15, 2017 (dataset 1) and three months from June 15, 2017 to September 15, 2017 (dataset 2). For Dataset 1, the first 40 days were used as the training set, the next 10 days as the validation set, and the remaining days as the test set. For Dataset 2, the first 60 days were used as the training set, the following 15 days as the validation set, and the remaining days as the test set. The prediction performance of ADBA model and five benchmark models on these two datasets are shown in Table 13, Table 14 and Fig. 18 respectively. It can be seen that all evaluation metrics results of ADBA model are significantly superior to those of other five benchmark models, indicating the ADBA hybrid model acquires the best prediction performance on different prediction time horizons.

4.11. Comparison between the proposed method and previous prediction methods

In this section, we compared the proposed method with some previous prediction approaches to demonstrate its superiority. Following the same experimental data setup as in Section 4.8, data from the Desert Gardens (Site 1) of DKASC was utilized for training and evaluating the overall performance of all models. All comparative models were implemented following their original network architectures and parameter settings as described in their respective papers. To further quantify the forecasting results using scientific values, following [14], the forecast skill score [52] was introduced to measure the ADBA and previous models. The common form of forecast skill can be given by

$$\text{forecast skill} = 1 - \frac{\text{error}_{proposed}}{\text{error}_{reference}} \tag{25}$$

where $\text{error}_{proposed}$ denotes the error of the evaluated model and $\text{error}_{reference}$ denotes the error of a reference model. In this study, the reference model is TCN [53], a common benchmark model with relatively good predictive performance. RMSE is the error metric used. According to Table 15, Dai et al. [54] proposed the BO-BCA-Seq2Seq

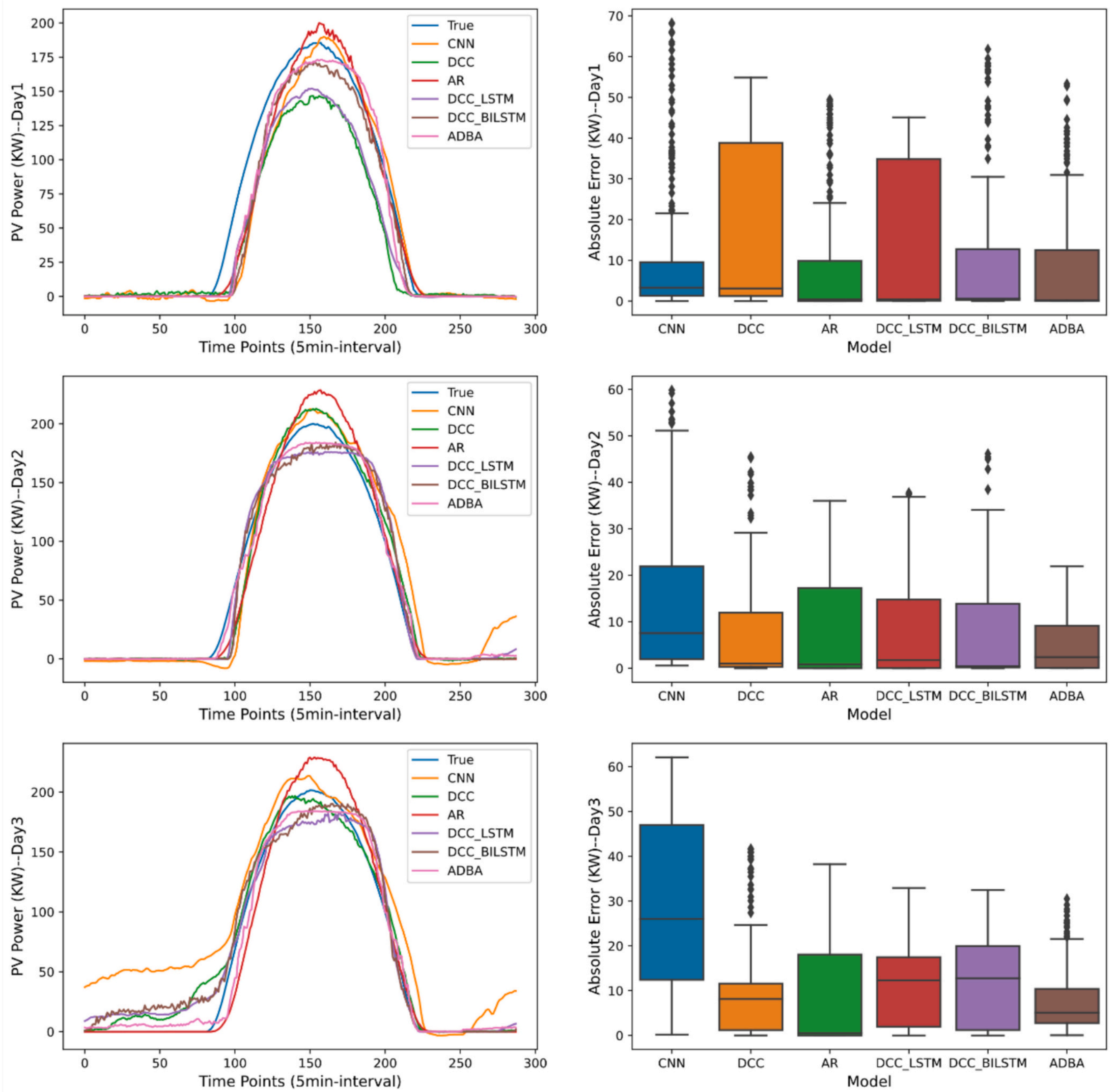

Fig. 15. Prediction performance of ADBA model and five benchmark models on Site 2 for three days.

model for short-term power load prediction by improving the sequence-to-sequence (Seq2Seq) model based on a bidirectional long-short term memory (Bi-LSTM) network. The forecast skill score of this model was 0.3648. Jiménez-Navarro et al. [55] designed PHILNet, a hierarchical prediction model combining Smooth Layer and LSTM, and the forecast skill score was 0.3524. Wang et al. [37] proposed a hybrid model combining LSTM and CNN (CLSTM for short), and its forecast skill score was 0.3375. LSTNet was designed by Lai et al. [32]. As a hybrid model combining CNN, RNN, and Autoregressive components, the forecast skill score of LSTNet was 0.2389. TCN is also a powerful prediction model based on causal convolution; the forecast skill score was 0 (because it is the reference model). Compared to these previous prediction methods, the proposed ADBA model demonstrated superior forecasting ability with the highest forecast skill score.

In addition, to further verify the training efficiency of the proposed method, we conducted a comparison of the training times of different models. As shown in Table 15, the total training time of the ADBA model proposed in this paper is 7.79 minutes, which is comparable to that of TCN (7.88 minutes) and significantly shorter than that of BO-BCA-Seq2Seq (8.94 minutes). This demonstrates the algorithm's competitive efficiency, thus enhancing the model's practical value.

5. Discussion and conclusions

This paper proposes an ADBA model for ultra-short-term PV power generation forecasting. The main conclusions can be summarized as follows:

(1) The ADBA is proposed based on the Attention mechanism and a parallel prediction architecture with linear AR and nonlinear DCC-BiLSTM components. The Attention mechanism optimizes multivariate

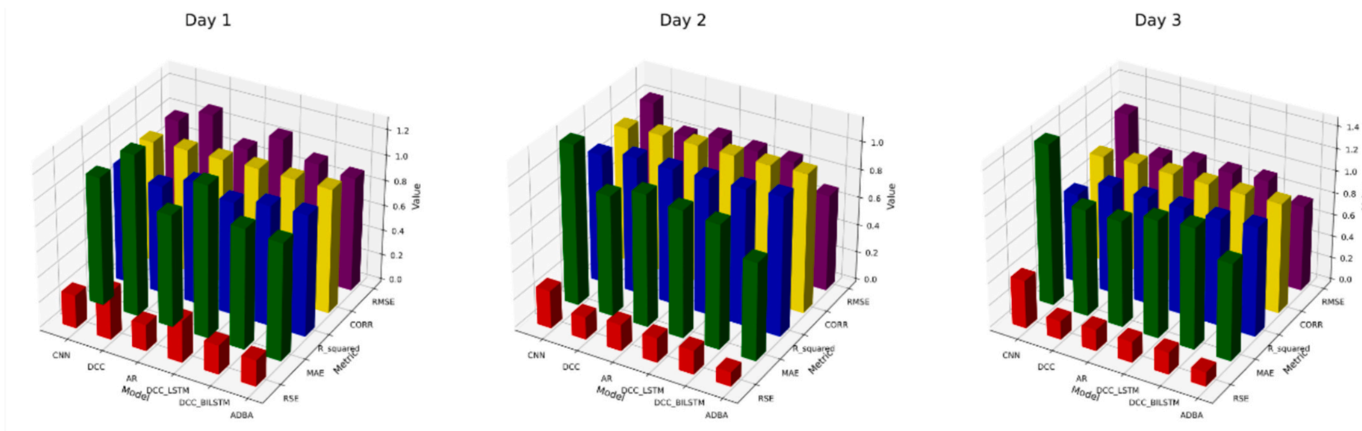

Fig. 16. The prediction performance metrics of ADBA model and five benchmark models on Site 2 for three days.

Table 7
The prediction performance metrics of ADBA model and five benchmark models on Site 2 for three days.

	Metrics	CNN	AR	DCC	DCC_LSTM	DCC_BiLSTM	ADBA
Day1	RSE	0.2628	0.2076	0.3757	0.3276	0.2305	**0.2069**
	CORR	0.9685	0.9803	0.9806	**0.9876**	0.982	0.9823
	MAE	10.3651	**7.6773**	18.5814	15.6562	8.8495	8.353
	R_squared	0.9309	0.9569	0.8588	0.8927	0.9469	**0.9572**
	RMSE	18.8642	14.8996	26.9683	23.5131	16.5442	**14.8497**
Day2	RSE	0.2706	0.1875	0.1596	0.1751	0.1696	**0.1027**
	CORR	0.9709	0.9874	0.9897	0.9846	0.9855	**0.9949**
	MAE	13.9891	8.7876	7.2427	8.0717	7.6522	**4.9768**
	R_squared	0.9268	0.9648	0.9745	0.9694	0.9712	**0.9895**
	RMSE	20.8513	14.4494	12.2974	13.4892	13.069	**7.9136**
Day3	RSE	0.4331	0.1918	0.1672	0.1875	0.1978	**0.129**
	CORR	0.9698	0.9876	0.9906	0.9869	0.9852	**0.992**
	MAE	27.9718	8.9647	9.1061	11.3877	12.168	**7.4192**
	R_squared	0.8124	0.9632	0.972	0.9648	0.9609	**0.9834**
	RMSE	33.8843	15.0032	13.0835	14.6709	15.4729	**10.0938**

Table 8
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 2.

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.2929	0.9586	14.2963	0.9142	21.5596
DCC	0.2362	0.9753	10.6161	0.9442	17.3918
AR	0.2039	0.9808	8.0917	0.9584	15.0091
DCC_LSTM	0.2024	0.9816	8.9331	0.959	14.8978
DCC_BiLSTM	0.1853	0.9827	7.8266	0.9657	13.6399
ADBA	**0.1533**	**0.9887**	**6.2873**	**0.9765**	**11.2833**

Table 9
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (spring).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.6869	0.8473	63.9703	0.5282	93.8034
DCC	0.5998	0.8219	50.9695	0.6403	81.9041
AR	0.4350	**0.9100**	34.3855	0.8108	59.4006
DCC_LSTM	0.4579	0.9092	37.5079	0.7904	62.5265
DCC_BiLSTM	0.4363	0.9038	36.6525	0.8096	59.5839
ADBA	**0.4345**	0.908	**33.1402**	**0.8112**	**59.3324**

Table 10
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (summer).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.2221	0.9776	38.0484	0.9507	47.3907
DCC	**0.135**	0.9917	**19.3296**	**0.9818**	**28.808**
AR	0.2038	0.9802	20.2378	0.9585	43.4842
DCC_LSTM	0.1596	0.9886	22.8679	0.9745	34.0482
DCC_BiLSTM	0.1751	**0.9927**	26.9739	0.9693	37.3743
ADBA	0.1505	0.9904	22.2953	0.9774	32.1117

Table 11
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (autumn).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.5452	0.8699	50.3598	0.7028	65.4131
DCC	0.3597	**0.9505**	32.3749	0.8706	43.1553
AR	0.3630	0.9363	31.9802	0.8682	43.5585
DCC_LSTM	0.4068	0.9303	37.4769	0.8345	48.8112
DCC_BiLSTM	0.4314	0.9276	40.2901	0.8139	51.7641
ADBA	**0.3573**	0.9439	**31.6429**	**0.8724**	**42.8671**

time series by considering the relative importance of different meteorological variables. The parallel architecture combines the complementary strengths of AR and DCC-BiLSTM components: the AR component excels at capturing linear patterns (showing better performance on data with stronger linear characteristics), while the DCC-BiLSTM component specializes in modeling complex nonlinear relationships (demonstrating superior performance on data with stronger nonlinear characteristics and have excellent capability in capturing both temporal and spatial features). This complementary design enables simultaneous modeling of both linear and nonlinear characteristics in historical power generation time series.

(2) The experimental results show that: ① The proposed ADBA

Table 12
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (winter).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.5459	0.8622	41.3671	0.7020	55.7032
DCC	0.4407	0.9278	34.1336	0.8058	44.9762
AR	0.2912	0.9578	20.9773	0.9152	29.7215
DCC_LSTM	0.3379	0.9523	25.8622	0.8858	34.4799
DCC_BiLSTM	0.3028	0.9670	22.4500	0.9083	30.9028
ADBA	**0.2588**	**0.9716**	**18.0472**	**0.9330**	**26.4118**

model achieves superior prediction performance on both Site 1 (with more linear characteristics, H = 0.268) and Site 2 (with stronger nonlinear patterns, H = 0.146), demonstrating its adaptability to different data patterns. ② Across four seasonal datasets, the ADBA model consistently outperforms benchmark models, with only slight inferiority to the DCC model in summer. ③ The model maintains excellent prediction accuracy across different time horizons (5, 10, and 15 minutes), indicating strong temporal robustness. ④ Comprehensive comparisons confirm the ADBA model's superior prediction accuracy and stability compared to existing methods.

While the ADBA model demonstrates significant improvements in prediction performance, future research could focus on: (1) further refining its computational efficiency to better support real-time applications, and (2) more feature selection methods of machine learning field such as Random Forest (RF) can be further combined to achieve more accurate feature selection result and ameliorate the quality of input.

CRediT authorship contribution statement

Zhengda Zhou: Writing – original draft, Software, Resources, Formal analysis, Conceptualization. **Yeming Dai:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Mingming Leng:** Writing – review & editing, Supervision.

Declaration of competing interest

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

Table 13
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (10 minutes).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.3925	0.9352	27.9402	0.8459	45.2836
DCC	0.2948	0.96	19.9746	0.9131	34.0078
AR	0.2971	0.9565	18.9006	0.9118	34.2707
DCC_LSTM	0.271	0.964	17.2547	0.9266	31.2648
DCC_BiLSTM	0.2766	0.964	17.0913	0.9235	31.9162
ADBA	**0.2279**	**0.9737**	**14.2003**	**0.948**	**26.2958**

Table 14
The overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (15 minutes).

	RSE	CORR	MAE	R_squared	RMSE
CNN	0.3729	0.9298	24.0531	0.8609	42.644
DCC	0.3121	0.9513	18.7212	0.9026	35.6921
AR	0.3453	0.9404	20.0115	0.8807	39.4887
DCC_LSTM	0.3008	0.9581	18.6332	0.9095	34.3967
DCC_BiLSTM	0.3038	0.9568	18.5311	0.9077	34.7395
ADBA	**0.2569**	**0.9673**	**15.8133**	**0.934**	**29.3771**

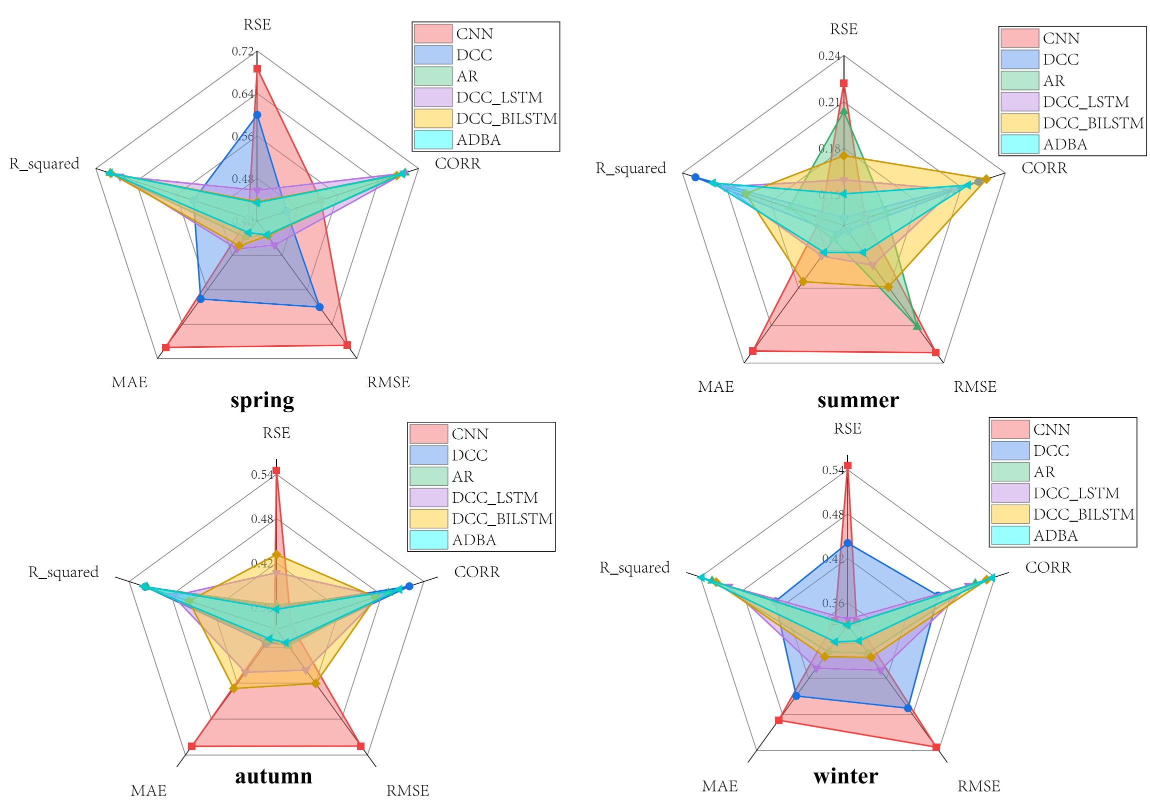

Fig. 17. Radar plot of the overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (four seasons).

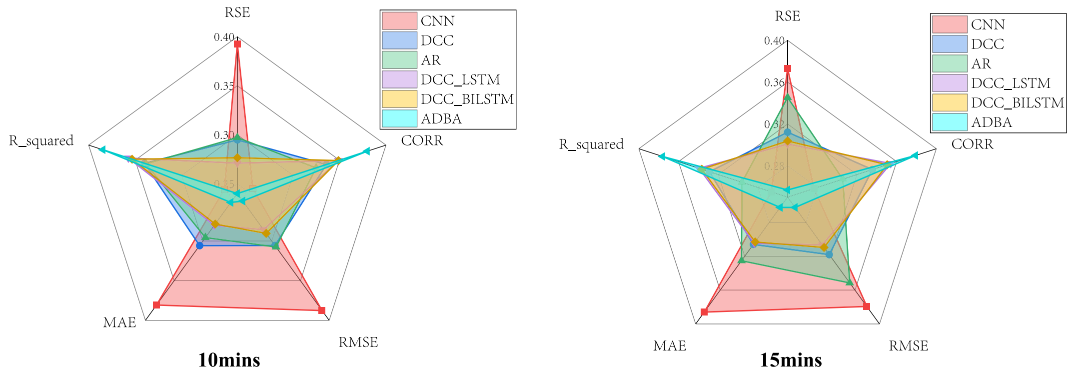

Fig. 18. Radar plot of the overall prediction performance metrics of ADBA model and five benchmark models on the entire test set of Site 1 (10 minutes and 15 minutes).

Table 15
Comparison between the proposed method and previous prediction methods.

Model	Study	Description	Forecast skill score	Training time cost (Epoch = 100, minutes)
BO-BCA-Seq2Seq	Dai et al.	A hybrid model based on the improved Seq2Seq, CNN, attention mechanism and Bayesian optimization	0.3648	8.94
PHILNet	Jiménez-Navarro et al.	A hierarchical prediction model combining Smooth Layer and LSTM	0.3524	3.84
CLSTM	Wang et al.	A hybrid model combining LSTM and CNN	0.3375	5.14
LSTNet	Lai et al.	A hybrid model combining CNN, RNN and Autoregressive.	0.2389	6.99
TCN (reference model)	Bai et al.	A powerful prediction model based on causal convolution to prevents interference from future information	0	7.88
ADBA	Ours	The proposed model	**0.3780**	7.79

appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.72371139), Humanities and Social Science Fund of Ministry of Education of China (No.20YJA630009) and Shandong Natural Science Foundation of China (No.ZR2022MG002).

Data availability

Data will be made available on request.

References

[1] Li H, Wang G, Duan H. A novel time-lagged logistic grey model and its application in forecasting energy production volume. Eng Appl Artif Intell 2024;127:107352.
[2] Ahmed R, Sreeram V, Mishra Y, et al. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. Renew Sust Energ Rev 2020;124:109792.
[3] Luo X, Zhang D, Zhu X. Combining transfer learning and constrained long short-term memory for power generation forecasting of newly-constructed photovoltaic plants. Renew Energy 2022;185:1062–77.
[4] Li Q, Zhang X, Ma T, et al. A multi-step ahead photovoltaic power prediction model based on similar day, enhanced colliding bodies optimization, variational mode decomposition, and deep extreme learning machine. Energy 2021;224:120094.
[5] Wang F, Xuan Z, Zhen Z, et al. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. Energy Convers Manag 2020;212:112766.
[6] Wang L, Tao R, Hu H, et al. Effective wind power prediction using novel deep learning network: stacked independently recurrent autoencoder. Renew Energy 2021;164:642–55.
[7] Ding S, Li R, Tao Z. A novel adaptive discrete grey model with time-varying parameters for long-term photovoltaic power generation forecasting. Energy Convers Manag 2021;227:113644.
[8] Liu J, Zang H, Cheng L, et al. A transformer-based multimodal-learning framework using sky images for ultra-short-term solar irradiance forecasting[J]. Appl Energy 2023;342:121160.
[9] Gao Y, Li P, Yang H, et al. A solar radiation intelligent forecasting framework based on feature selection and multivariable fuzzy time series. Eng Appl Artif Intell 2023;126:106986.
[10] Li X, Law R. Forecasting tourism demand with decomposed search cycles[J]. J Travel Res 2020;59(1):52–68.
[11] Chu Y, Urquhart B, Gohari SMI, et al. Short-term reforecasting of power output from a 48 MWe solar PV plant. Sol Energy 2015;112:68–77.
[12] López JC, Rider MJ, Wu Q. Parsimonious short-term load forecasting for optimal operation planning of electrical distribution systems. IEEE Trans Power Syst 2018;34(2):1427–37.
[13] Li H, Wang G, Duan H. A novel time-lagged logistic grey model and its application in forecasting energy production volume. Eng Appl Artif Intell 2024;127:107352.
[14] Tang Y, Yang K, Zhang S, et al. Photovoltaic power forecasting: a hybrid deep learning model incorporating transfer learning strategy. Renew Sust Energ Rev 2022;162:112473.
[15] Liu Y, Wang J. Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. Appl Energy 2022;312:118729.
[16] Yin Y, Yu Z, Gou X, et al. Photovoltaic power prediction model based on empirical mode decomposition-long-short memory neural network[C]//2021 international conference on intelligent computing, automation and systems (ICICAS). IEEE 2021:349–53.
[17] Benali L, Notton G, Fouilloy A, et al. Solar radiation forecasting using artificial neural network and random forest methods: application to normal beam, horizontal diffuse and global components. Renew Energy 2019;132:871–84.
[18] Kaytez F. A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. Energy 2020;197:117200.
[19] Voyant C, Motte F, Notton G, et al. Prediction intervals for global solar irradiation forecasting using regression trees methods. Renew Energy 2018;126:332–40.
[20] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673–81.
[21] Chen X, Gupta L. Training LSTMS with circular-shift epochs for accurate event forecasting in imbalanced time series. Expert Syst Appl 2024;238:121701.
[22] Zang H, Cheng L, Ding T, et al. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. IET Gener Transm Distrib 2018;12(20):4557–67.
[23] Zhao Z, Yun S, Jia L, et al. Hybrid VMD-CNN-GRU-based model for short-term forecasting of wind power considering spatio-temporal features. Eng Appl Artif Intell 2023;121:105982.
[24] Niu T, Li J, Wei W, et al. A hybrid deep learning framework integrating feature selection and transfer learning for multi-step global horizontal irradiation forecasting. Appl Energy 2022;326:119964.
[25] Zheng W, Huang L, Lin Z. Multi-attraction, hourly tourism demand forecasting. Ann Tour Res 2021;90:103271.
[26] Zhang C, Ma H, Hua L, et al. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. Energy 2022;254:124250.

[27] Ji J, Tan Z, Wong M, et al. Heat pipe-constrained IoT device layout via multiobjective differential evolution. IEEE Internet Things J 2024. https://doi.org/10.1109/JIOT.2024.3498445. online.
[28] Ajith M, Martínez-Ramón M. Deep learning based solar radiation micro forecast by fusion of infrared cloud images and radiation data. Appl Energy 2021;294:117014.
[29] Ghimire S, Deo RC, Raj N, et al. Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms. Appl Energy 2019;253:113541.
[30] Zhang C, Peng T, Nazir MS. A novel integrated photovoltaic power forecasting model based on variational mode decomposition and CNN-BiGRU considering meteorological variables. Electr Power Syst Res 2022;213:108796.
[31] Feng R, Chen M, Song Y. Learning traffic as videos: short-term traffic flow prediction using mixed-pointwise convolution and channel attention mechanism. Expert Syst Appl 2024;240:122468.
[32] Lai G, Chang WC, Yang Y, et al. Modeling long-and short-term temporal patterns with deep neural networks[C]//The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. 2018. p. 95–104.
[33] Liu K, Ruan J, Zhao X, Liu G. Short-term load forecasting method based on sparrow search optimized attention-GRU. Proc CSU-EPSA 2022;34:1–9.
[34] Zhang H, Gao J, Ye X, et al. Power system operation modes identification based on data-driven and deep learning[C]//2023 panda forum on power and energy (PandaFPE). IEEE 2023:1678–84.
[35] Tong C, Zhang L, Li H, et al. Attention-based temporal–spatial convolutional network for ultra-short-term load forecasting. Electr Power Syst Res 2023;220:109329.
[36] Li Z, Gao Z, Zhang X, et al. Time-aware personalized graph convolutional network for multivariate time series forecasting. Expert Syst Appl 2024;240:122471.
[37] Wang K, Qi X, Liu H. Photovoltaic power forecasting based on LSTM-convolutional network. Energy 2019;189:116225.
[38] Fang X, Gong G, Li G, et al. A hybrid deep transfer learning strategy for short-term cross-building energy prediction. Energy 2021;215:119208.
[39] Li J, Lin T, Du H, et al. A wind power prediction model based on optimized N-BEATS network with multivariate inputs[C]//2023 IEEE Power & Energy Society General Meeting (PESGM). IEEE 2023:1–5.
[40] Meng A, Xie Z, Luo J, et al. An adaptive variational mode decomposition for wind power prediction using convolutional block attention deep learning network. Energy 2023;282:128945.
[41] Kumar I, Tripathi BK, Singh A. Attention-based LSTM network-assisted time series forecasting models for petroleum production. Eng Appl Artif Intell 2023;123:106440.
[42] Kulshrestha A, Krishnaswamy V, Sharma M. Bayesian BILSTM approach for tourism demand forecasting. Ann Tour Res 2020;83:102925.
[43] Zhang C, Ma H, Hua L, et al. An evolutionary deep learning model based on TVFEMD, improved sine cosine algorithm, CNN and BiLSTM for wind speed prediction. Energy 2022;254:124250.
[44] Liu S, Huang Q, Li M, et al. A new LASSO-BiLSTM-based ensemble learning approach for exchange rate forecasting. Eng Appl Artif Intell 2024;127:107305.
[45] Box GEP, Jenkins GM. Box and Jenkins: Time series analysis, forecasting and control[M]//in: A very British affair: Six Britons and the development of time series analysis during the 20th century. London: Palgrave Macmillan UK; 2013. p. 161–215.
[46] Hyndman RJ. Forecasting: Principles and practice[M]. OTexts; 2018.
[47] Li C, Zheng W, Ge P. Tourism demand forecasting with spatiotemporal features. Ann Tour Res 2022;94:103384.
[48] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. p. 770–8.
[49] Lin TY, Dollár P, Girshick R, et al. Feature pyramid networks for object detection [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. p. 2117–25.
[50] Ronneberger, O., Fischer, P., Brox, T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III. Springer Int Pub, 2015: 234–241.
[51] Meng A, Chen S, Ou Z, et al. A novel few-shot learning approach for wind power prediction applying secondary evolutionary generative adversarial network. Energy 2022;261:125276.
[52] Yang D. A guideline to solar forecasting research practice: reproducible, operational, probabilistic or physically-based, ensemble, and skill (ROPES). J Renew Sustain Energy 2019;11(2):023704.
[53] Bai S, Kolter JZ, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. arXiv preprint arXiv:1803.01271. 2018.
[54] Dai Y, Yang X, Leng M. Optimized Seq2Seq model based on multiple methods for short-term power load forecasting. Appl Soft Comput 2023;142:110335.
[55] Jiménez-Navarro MJ, Martínez-Ballesteros M, Martínez-Álvarez F, et al. PHILNet: a novel efficient approach for time series forecasting using deep learning. Inf Sci 2023;632:815–32.