Applied Energy 251 (2019) 113315

Contents lists available at ScienceDirect

Applied Energy

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

A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network

Kejun Wang, Xiaoxia Qi, Hongda Liu*

College of Automation, Harbin Engineering University, Harbin 150001, China

HIGHLIGHTS

- Three deep learning neural networks are proposed for photovoltaic power forecasting.
- The influence of input sequence length on prediction accuracy is considered.
- Provide suggestions for choosing the most suitable network in practical application.
- The better effect of deep learning on photovoltaic power forecasting is demonstrated.
- The stability and robustness of the proposed model are proved.

ARTICLE INFO

Keywords:
Deep learning
Convolutional neural network
Long short-term memory
Photovoltaic power prediction

ABSTRACT

Accurate photovoltaic power forecasting is of great help to the operation of photovoltaic power generation system. However, due to the instability, intermittence, and randomness of solar energy, accurate prediction of photovoltaic power forecasting becomes very difficult. In this paper, the convolutional neural network, long short-term memory network, and hybrid model based on convolutional neural network and long short-term memory network models were proposed, and are applied to the obtained data in DKASC, Alice Springs photovoltaic system. The mean absolute percentage error, root mean square error, and mean absolute error indicators are used to evaluate the performance of the prediction model in this paper. The results showed that when the input sequence is increased, the accuracy of the model is also improved, and the prediction effect of the hybrid model is the best, followed by that of convolutional neural network. While long short-term memory network had the worst prediction effect, the training time was the shortest. However, not the longer the input sequence is, the better the prediction will be. This may be related to the characteristics of the time series itself. The results of the deep learning model proposed in this paper on photovoltaic power prediction also indicate that deep learning is very helpful for improving the accuracy of PV power prediction.

1. Introduction

As one of the most popular renewable energy sources, solar energy has the advantages of abundant resources, no pollution, free use, and no transportation. At present, solar power generation is mainly divided into solar thermal power generation and photovoltaic (PV) power generation, and PV power generation is growing at a relatively fast rate every year. Due to the randomness of light and the periodicity of day and night, PV power plants have natural uncontrollability and are typical fluctuating and intermittent power sources [1]. The output of PV power generation system is largely affected by weather, climate and other factors. These characteristics lead to new challenges to the power system after a high proportion of PV access, such as increasing the difficulty and complexity of grid scheduling [2]. Power forecasting for PV power generation has become one of the key basic technologies for improving the quality of operational scheduling and reducing spare capacity reserves.

According to the classification of prediction process, PV power prediction can be divided into direct prediction and indirect prediction; According to the classification of the spatial scale of prediction, it can be divided into single field prediction and regional prediction. According to the classification of the time scale of prediction, it can be divided into ultra-short-term forecast, short term forecast, medium term forecast and long term forecast. According to the classification of the prediction form, it can be divided into point prediction, interval prediction and probability prediction. According to the classification of

* Corresponding author.
E-mail address: liuhd405@hrbeu.edu.cn (H. Liu).

https://doi.org/10.1016/j.apenergy.2019.113315
Received 1 November 2018; Received in revised form 7 May 2019; Accepted 12 May 2019
Available online 16 May 2019
0306-2619/

Nomenclature

ANN	artificial neural network
ARIMA	auto regressive integrated moving average
BP	back propagation
CNN	convolutional neural network
DBN	deep belief network
DNNs	deep neural networks
ELM	extreme learning machine
FC	fully connected
GRNN	generalized regression neural network
LSTM	long short term memory
MAE	mean absolute error
MAPE	mean absolute percentage error
MEPSO	macedonian system operator
MRE	missing rate error
NWP	numerical weather prediction
PV	photovoltaic
RBFNN	radial basis function neural network
RMSE	root mean square error
RNN	recurrent neural network
SDAE	stacked denoising auto-encoder
SVR	support vector regression

the different prediction methods, it can be divided into physical method, statistical method and machine learning method. However, for any kind of classification, prediction research can be carried out by different prediction methods such as physical method, statistical method and machine learning method [3].

The energy conversion process of PV power generation is affected by the correlation between solar irradiance and uniformity (ie, shadow), pollution, aging, battery temperature, solar incident angle, and load conditions. The physical method is a mathematical model established according to the principle of PV power generation, using solar radiation, temperature, humidity, cloud volume, air pressure and wind speed obtained by numerical weather prediction (NWP), combined with PV system installation angle, PV array conversion efficiency, battery conditions and other parameters to establish the physical model and then directly calculates the PV power. The physical prediction model does not require historical data, but relies on detailed station geographical information, accurate meteorological data and complete PV battery information. The prediction accuracy of the physical method is strongly dependent on the accuracy of the NWP information, but currently it encounters bottlenecks in improving the accuracy of NWP [4]. In addition, the parameters provided by PV manufacturers are often missing and of limited accuracy; Due to the relation of cognition degree, the established physical correlation model has certain errors, and the model needs to rely on empirical parameters (threshold), while the empirical parameters are different between different regions, resulting in poor local anti-interference ability and weak robustness [5].

Common statistical prediction methods include time series method [6], regression analysis method [7], gray theory [8], fuzzy theory [9] and spatio-temporal correlation method [10]. The statistical method is to establish the correlation mapping relationship between input-output data (i.e., data model) by means of curve fitting, parameter estimation and correlation analysis of the processed historical data such as solar radiation and PV power generation output, so as to realize the prediction of future PV power generation output. Different from the physical method, statistical modeling does not require a clear and complete understanding of the complex photoelectric conversion relationship of PV system in advance, but only a partial understanding and realization through various data analysis techniques. Therefore, compared with the physical method, it has the advantages of simple modeling and strong universality between different regions. However, the precondition for the implementation of statistical methods is that they need to have a large number of correct historical data after processing, and there are difficulties in data acquisition and calculation processing during the implementation [5]. A lot of numerical calculation is needed in the prediction process, and it usually takes a long time to predict by ordinary computer, so it is difficult to meet the requirement of ultra-short-term prediction (especially minute level) on the prediction speed. In addition, the prediction quality of statistical method is strongly related to the quality of data, and the reserve of historical data. The data screening and the elimination effect of false data have a great impact on the prediction accuracy. The prediction accuracy usually depends on the calculation of higher dimensions to ensure the effect, which increases the calculation and slows down the prediction speed.

Machine learning has the ability to efficiently extract high-dimensional complex nonlinear features and map them directly to the output. The machine learning-based prediction method takes advantage of this and has become one of the most commonly used methods for predicting time series [11,12]. In [13,14], a solar power prediction model based on ANN was proposed, and meteorological data (such as irradiance, temperature, humidity, wind speed and air pressure, etc.) were selected as the input of the model in the research. The results show that the prediction performance of the ANN-based prediction model is better than some typical physical or statistical (multiple linear regression (statistical), analytical and one-diode models (physical), etc.) prediction models. In order to further explore the prediction model with better prediction performance, Yadav AK et al. [15] developed an ANN-based irradiance prediction model using the artificial neural network fitting tool (nftool), and used the rapid miner technique to select the most relevant input variables and compare them with a variety of different artificial neural network prediction models (such as RBFNN, GRNN). The results shown that the ANN model developed using the artificial neural network fitting tool has the best predictive irradiance effect, and the maximum input (such as the highest temperature, the lowest temperature, the average temperature, irradiance, etc.) have the best effect, and the value of MAPE can reach 6.89%. Cervone G et al. [16] proposed a hybrid method based on neural network and analog ensemble for short-term photovoltaic power prediction for photovoltaic fields in three regions of Italy. The proposed method has achieved better performance on both deterministic and probabilistic PV power prediction (the RMSE values of the deterministic prediction indicators in the three regions reached 8.09%, 7.39%, 8.66%, respectively, and the MRE values of the probabilistic evaluation indicators reached −1.85%, 0.38%, −1.53%, respectively). Almonacid F et al. [17] proposed a short-term prediction method of photovoltaic power based on dynamic neural network, which can be used to predict the power output of photovoltaic system one hour in advance with acceptable accuracy (the RMSE is up to 3.38%).

With the development of technology in earth sciences and atmospheric sciences, and the open source of internet for commercial/civilian meteorological data and NWP data, the forecastable data has increased relative to the past. Currently available data such as numerical weather prediction, data from low-speed communication networks, and satellite monitoring data. Traditional neural networks (shallow networks) in the new situation of facing more input variables, shallow network cognitive ability is limited, in order to recognize more complex relationships between dimensional input and output, only rely on increase the input layer node, the number of hidden layers, and the number of hidden layer nodes. However, in some cases, the effect is not significantly improved compared to the physical method using the better coefficient [18]. Furthermore, for traditional ANN networks, excessive input data, hidden layer numbers and hidden layer nodes are likely to cause over-fitting, gradient disappearance and explosion of network training. A more efficient learning method is also needed for this [19].

Deep neural networks (DNNs) have higher feature extraction capabilities than shallow neural networks and can significantly improve the problem of gradient disappearance of shallow neural network [20]. Common deep neural networks mainly include Convolutional Neural Network (CNN) [21], Deep Belief Network (DBN) [22], Stacked Denoising Automatic Encoder (SDAE) [23], and Long Short Term Memory (LSTM) [24]. DNNs have achieved remarkable results in pattern recognition, speech recognition, and language translation. In 2016, H.Z. Wang et al. [25] used DNNs for time series prediction for the first time. The deep learning-based predictive model can take into account the temporal and nonlinear characteristics of time series data, and can find complex data associations from a larger amount of data. In theory, the prediction model based on DNNs should have better performance and robustness than the traditional shallow network model.

So far, predictive models based on deep learning have achieved certain results in the prediction of wind speed, wind power, load, and solar irradiance. Wan J et al. [26] proposed a prediction method based on DBN for wind speed prediction. The research results show that compared with the shallow network of SVR, single hidden layer neural network and multiple hidden layer neural network, the prediction model based on DBN has better network performance and higher prediction accuracy. Dedinec A et al. [27] used the DBN model to predict the power load of Macedonian. By comparing the proposed method with multi-layer feedforward neural network prediction data and MEPSO prediction data, the results show that the new model has better prediction accuracy (MAPE values are reduced by 8.6% and 21%, respectively). Liu H et al. [28], proposed a new hybrid deep learning model for wind speed prediction. Wavelet transform is used to process the network input data, divided the data into several sub-layers. LSTM is used to predict the low frequency sublayer and Elman neural network is used to predict the high frequency sublayer. Compared with the prediction results of 11 prediction models based on ARIMA, BP and

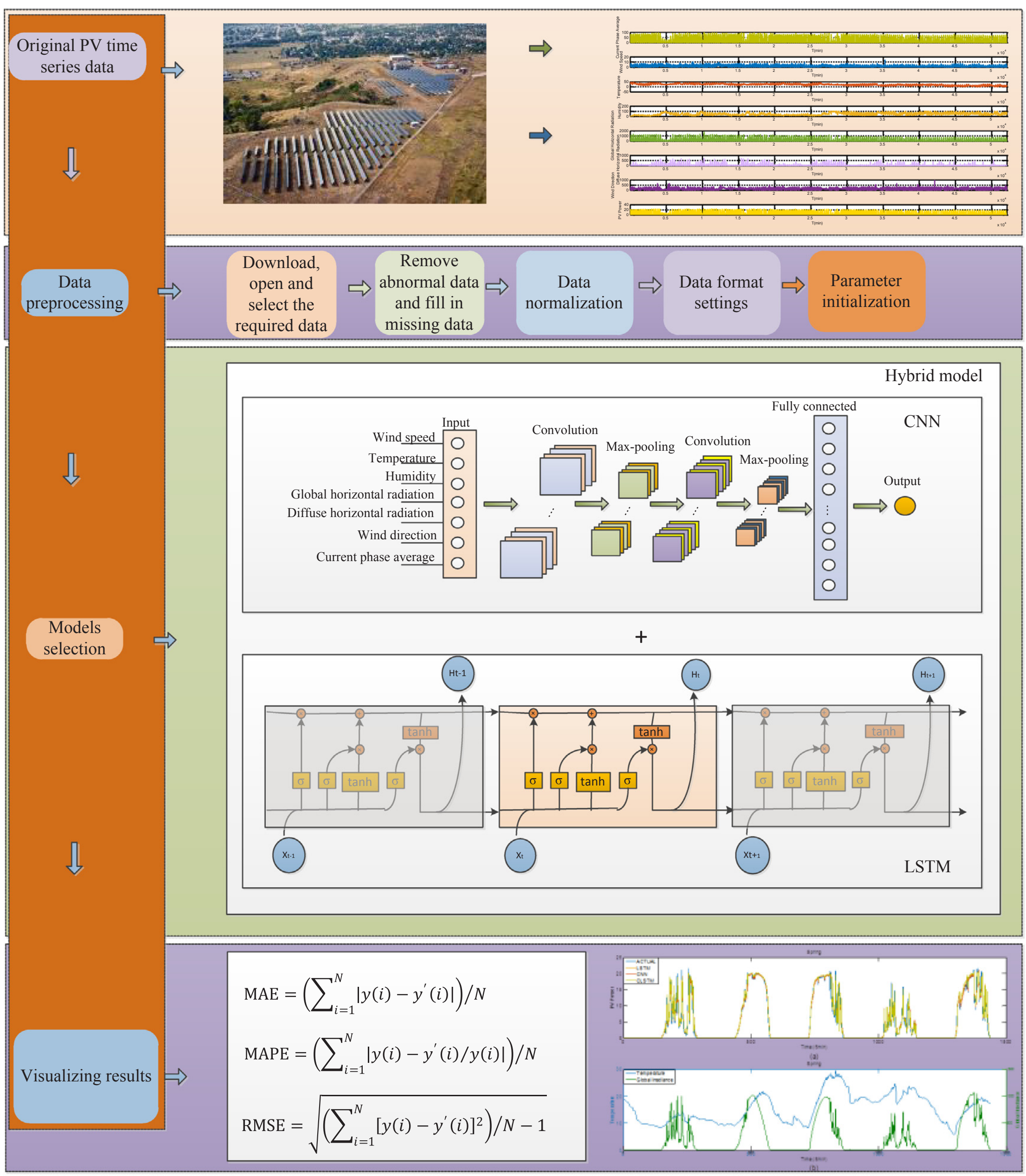


Fig. 1. The structure of the proposed model.

GRNN, etc., the mixed model has satisfactory performance (MAPE, MAE and RMSE values are all lower than the 11 prediction models). Liu H et al. [29], proposed a hybrid deep learning model based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM to improve the accuracy and robustness of the wind speed prediction model and compared with 8 models (MAPE, MAE and RMSE values all decreased to a large extent). The experimental results show that the proposed model has the best prediction performance and is more effective in extracting trend information. Wang H et al. [30], proposed CNN model for probabilistic wind power prediction and compared with continuous model, SVR and BP shallow model. The results show that the proposed method can better learn the uncertainty of data and verify the competitive performance of this method.

Compared to wind speed/wind power prediction, PV power prediction has more influencing factors and available data. Qing X et al. [31], selected the weather forecast data and meteorological data as the input of the prediction network, and the LSTM model was used to predict the solar irradiance of the day. And by comparing the three prediction algorithms with persistence method, least squares regression and back propagation, the RMSE values are reduced by 63.7%, 66.9% and 42.9%, respectively, and show less overfitting and generalization ability. The prediction of PV power is an indirect prediction using a mathematical model between irradiance and PV power established by physical methods. However, the prediction results of PV power are greatly affected by the accuracy of physical model modeling, and at the same time face a large number of calculation formulas and calculation processes. This is because the PV output power is related to the three irradiances of diffuse horizontal irradiance, direct normal irradiance, and extraterrestrial solar irradiance [32], and also to factors such as PV plate temperature [33] and atmospheric humidity [13]. Direct prediction of PV power using deep networks requires that the network can simultaneously describe these associations, so there are few related studies. This requires the network to have more powerful capabilities in identifying data uncertainty characteristics than the [31] network. Wang H et al. [34], proposed a hybrid prediction model combined wavelet transform and CNN, which realizes direct prediction of deterministic and probabilistic PV power. Since CNN has special requirements for the input data format, the "one-dimensional data to two-dimensional image layer" and "two-dimensional image to one-dimensional data layer" links are added to the model. Although it can meet the requirements and achieve good performance, in the convolution operation, due to the local role of the convolution kernel, some false information may be caused and the data conversion increases the complexity of the model.

In this paper, three PV power direct prediction models are proposed, which are: one-dimensional convolutional neural network model, long-short-term neural network model and "CNN + LTSM" hybrid model, which omits the data conversion link, and predictive performance was compared to find the most appropriate deep learning model for PV power prediction.

The main research contents of this paper are as follows:

(1) Three PV power direct prediction models based on deep learning neural network (CNN, LSTM and hybrid models based on CNN and LSTM) were designed and proposed, and the prediction performance of each model was compared.
(2) The relationship between the length of different input sequences and the prediction accuracy of the model is discussed, and the length of the input sequence which is more suitable for the three prediction models is given. This helps to select the most appropriate forecasting model according to the degree of accumulation of different historical data in the project implementation.
(3) The experimental results show that the deep learning network has a good effect on the prediction of photovoltaic power generation and the stability and robustness of the model are high. The training time for each model presented in this paper provides recommendations for developers to consider the training time selection of the model.

The main structure of the paper is described as follows: Section 2 introduces the structure of the proposed model; Section 3 describes the experimental methods that were proposed in this paper; In Section 4 shows the performance indexes in this paper. Section 5 shows the case study results and provides an analysis and comparison of experimental results; Section 6 gives the conclusion of the article.

2. The structure of the proposed model

The whole structure of the mentioned model in this paper is described in Fig. 1. For a deeper understanding of the proposed model, a more detailed description is given below:

1. Data preprocessing

Before the data is transmitted to the network, it needs to be preprocessed, including removing abnormal data (in extreme circumstances) and filling in missing data (such as equipment failure). A large amount of data has a better effect on the deep learning network, so the data augmentation is used. Use data normalization to unify the data into the same unit of measure. Because deep convolutional neural networks and LSTM have special requirements for the format of the input data, it is necessary to turn the data into the appropriate format. This section also includes initializing the network.

2. Convolutional neural network

Since the data obtained is time series data, so the 1D CNN network include convolutional layer, pooling layer, and fully connected layer is proposed. The convolution and pooling layers are used to obtain depth information of time series.

3. Long short term memory neural network

LSTM neural network is a very popular network, which belongs to a kind of RNN networks. Its greatest feature is the introduction of the concept of the "gate", and the details are introduced in Section 3.2.

4. The hybrid CNN and LSTM neural network

CNN and LSTM have good results in small time series prediction tasks. CNN can extract spatial features and LSTM can extract temporal features, therefore a model that mixes CNN with LSTM is introduced and compared separately.

3. The description of the experimental methods

3.1. Convolutional networks and hierarchical structure of CNN

CNNs are a specialized kind of neural network for processing data that has a known, grid-like topology. Examples include time-series data, which can be thought of as a 1D grid taking samples at regular time intervals, and image data, which can be thought of as a 2D grid of pixels. CNNs have been tremendously successful in practical applications. The name "convolutional neural network" indicates that the network employs a mathematical operation called convolution. Convolution is a specialized kind of linear operation. CNNs are simply neural networks that use convolution in place of general matrix multiplication in at least one of their layers [35].

With the development of CNNs, many variants of convolutional network structures have emerged, but most of their basic structures are similar, including the input layer, convolutional layer, excitation layer, pooling layer, fully connected layer, and output layer

3.1.1. Input layer

The input layer is mainly for the operation of input data. Before the data is transmitted to the input layer, it needs to be processed. There are mainly three common methods, including de-averaging (centering all dimensions of the input data to 0), normalization (normalize the data amplitude to the same range), PCA (reduce the data dimension), etc. For a convolutional neural network, there are multiple formats of input data, including 1-D, 2-D, and even n-D, and these data usually consist of 1 to n channels. The examples of data formats with different dimensionalities and number of channels are shown in Table 1 [36]. In this paper, the 1-D input data format was used for the forecasting model that proposed in this paper.

3.1.2. Convolutional layer

The convolutional layer is the basic operation in the convolutional neural network, and it's a kind of local operation. It obtains the local information of the image through a certain size convolution kernel acting on the partial image area. Local connections and weight sharing are two important features of convolutional neural networks. The specific steps of the convolution operation can be described in Fig. 2.

3.1.3. Excitation layer

The activation function layer is also called the non-linearity mapping layer, and the purpose is to increase the expressiveness (non-linearity) of the entire network. Table 2 describes the most commonly used activation functions in current deep convolutional neural networks [37].

3.1.4. Pooling layer

Pooling layer is actually a down-sampling operation, sandwiched between successive convolution layers. It can compress the amount of data and parameters, reducing over-fitting. In this section, some of the pooling methods recently used in CNNs are described. That is Lp pooling (based on complex cell operating mechanisms, inspired by biology), average-pooling (take the average value of the pooled kernel coverage area as pooling result), max-pooling (take the max value of the pooled kernel coverage area as pooling result), mixed pooling (combine average- pooling with max-pooling, inspired by random dropout and drop-connect), spectral pooling (perform a DFT transform on the feature, then cut out the required spectrum in the frequency domain, and use IDFT to return the airspace), and stochastic-pooling (inspired by dropout, in a global sense, it is similar to the average-pooling, and in a local sense, it obey the maxi-pooling criteria).

3.1.5. Excitation layer

The fully connected layer functions as a "classifier" in the entire convolutional neural network, usually at the end of the convolutional neural network. All neurons between two layers have the weight to reconnect.

3.2. Long-Short-Term memory neural network

The LSTM network is a type of recurrent neural network (RNN). Like all recurrent neural networks, LSTM can calculate everything that a traditional computer can calculate with enough network elements.

3.2.1. The description of LSTM

Traditional feed-forward neural networks only accept information from the input node. It can only operate on the input space, and does not "remember" input to different time series. In feed-forward neural networks, information can only flow from the input layer to the hidden layer and then to the output layer. The biggest difference between RNN and feed-forward neural network is that it not only can operate on the input space but also can operate on the internal state space [38]. Its unfold structure is shown in Fig. 3.

In practical applications, the aforementioned RNN model has problems of gradient disappearance and gradient explosion. In order to solve this problem, Sepp Hochreiter and Jürgen Schmidhuber proposed the LSTM algorithm [39]. LSTM introduces the concept of input and output gates. The basic unit of LSTM is called a memory cell. Later, someone further expanded LSTM and introduced the forget gate. The specific information dissemination of LSTM is shown in Fig. 4 [40].

The LSTM information dissemination process can be explained as follows, and in Fig. 4, the f_t, i_t, o_t, g_t, represent the output value of the forget gate, input gate, output gate, and update gate respectively. The four gates receive the LSTM output value h_{t-1} at a former time step $t-1$ and the input data x_t at a present time step$t$ as the input.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$g_t = tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \tag{3}$$

$$c_t = f_t * c_{t-1} + i_t * g_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(c_t) \tag{6}$$

where $W_{f,i,g,o}$ and $b_{f,i,g,o}$ represent the weight matrices and bias vectors of the abovementioned gates. c_t represents the memory cell. σ and *tanh* represent the *sigmoid* and *hyperbolic tangent* activations.

4. Performance indexes

In order to assess the forecasting performance of the proposed models in this paper, three accuracy estimating metrics were selected in the prediction experiments. They are MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error), and RMSE (Root Mean Square Error), and can be expressed as follows:

$$\text{MAE} = \left(\sum_{i=1}^{N} |y(i) - y^{'}(i)|\right)/N \tag{7}$$

$$\text{MAPE} = \left(\sum_{i=1}^{N} |y(i) - y^{'}(i)/y(i)|\right)/N \tag{8}$$

$$\text{RMSE} = \sqrt{\left(\sum_{i=1}^{N} [y(i) - y^{'}(i)]^2\right)/N - 1} \tag{9}$$

$$\text{Loss} = \sum_{i=1}^{N} [y(i) - y^{'}(i)]^2/N \tag{10}$$

where $y(i)$ is the actual PV power series, $y^{'}(i)$ is the forecast values, and N is the number of$y(i)$.Loss is the loss function chosen in this paper.

5. Case study

5.1. The description of the experimental data

In this paper, the 1B DKASC, Alice Springs PV system data was selected in the following case research. The specific information of this system is shown in Table.3. From [41], the required data for the model is downloaded. The data in four years (2014–2017) were selected for this experiment. The data resolution is 5-min. Fig. 5 shows the historical data of consecutive days from 2014 to 2017. The picture depict

Table 1
The examples of data formats with different dimensionalities and number of channels.

	Single channel	Multi-channel
1-D	Audio waveform	Skeleton animation data
2-D	Audio data that has been preprocessed with a Fourier transform	Color image data
3-D	Volumetric data: such as CT scans.	Color video data

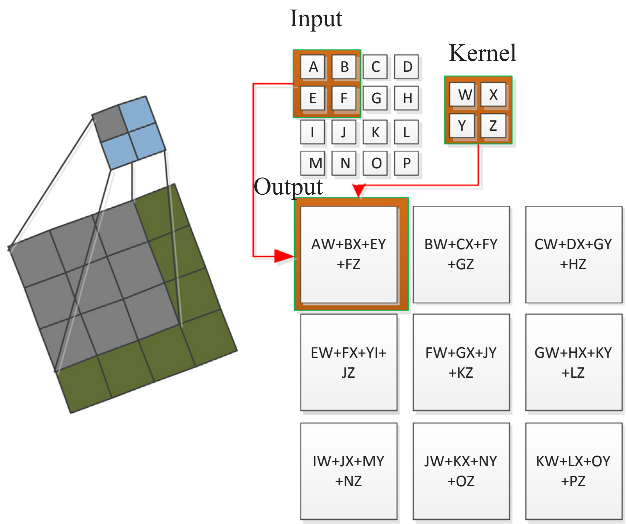


Fig. 2. The specific steps of the convolution operation.

historical data of different years, but are actually the same date. Without loss of generality, each graph describes data for a continuous week, where the horizontal axis represents time and the vertical axis represents power value. From any of the figures, it can be seen that the power in the noon period is the maximum, which is relatively small in the morning and afternoon, while tends to 0 in the evening. PV power between adjacent dates is affected by various factors and varies greatly without strong regularity. According to the comparison of four figures,

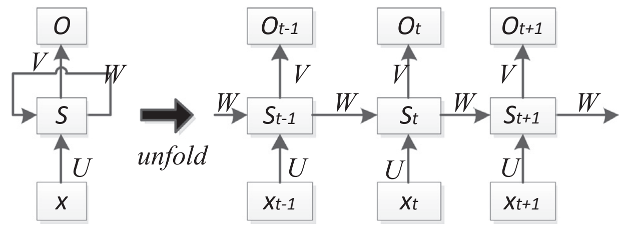
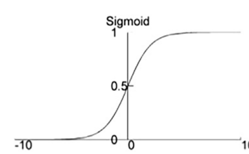
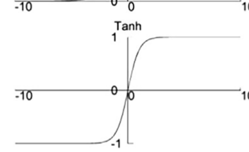
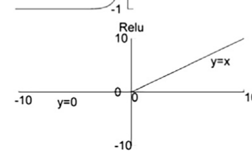
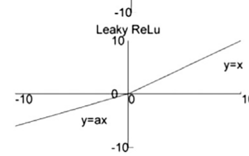
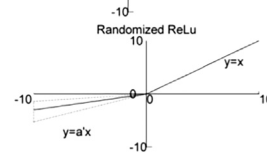
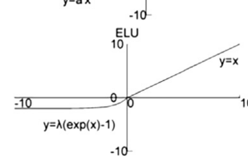


Fig. 3. The unfold structure of RNN.

due to the influence of natural environment factors, the power difference between the same dates in different years is also significant and hard to find the rule, that is, there is also high uncertainty. The map of the system is shown in Fig. 6 and the solar power installations in Alice Springs, Central Australia contains 38 sites, among which 1site and 9site are divided into A and B groups. The data used in this paper is from 1B. The detailed information of all solar power installations can be found in [42].

The downloaded data mainly includes current phase average (A), active power (kW), wind speed (m/s), weather temperature celsius(°C), weather relative humidity (%), global horizontal radiation (w/$m^2 \times sr$), diffuse horizontal radiation (w/$m^2 \times sr$), wind direction (Â°), etc. The normalization method is used to divide data into the same dimension, and the "data augmentation" technology is adopted to increase the number of samples (such as the wind direction sine value and cosine value is used). The data division criterion in this article is shown in Fig. 7.

Table 2
Activation functions in current deep convolutional neural networks.

	Activation function	Function shape
Sigmoid	$\sigma(x) = \frac{1}{1 + exp(-x)}$	
Tanh(x)	$\tanh(x) = 2\sigma(2x) - 1$	
ReLu	$\mathrm{ReLu}(x) = \max\{0, x\} = \begin{cases} x & ifx \geq 0 \\ 0 & ifx < 0 \end{cases}$	
Leaky ReLu or Parametric ReLu	$\mathrm{LeakyReLu}(x) = \begin{cases} \mathrm{x} & ifx \geq 0 \\ a{\cdot}x & ifx < 0 \end{cases}$	
Randomized ReLu	$\mathrm{RandomizedReLu}(x) = \begin{cases} x & ifx \geq 0 \\ a^{'}{\cdot}x & ifx < 0 \end{cases}$	
ELU	$\mathrm{ELU}(x) = \begin{cases} \mathrm{x} & ifx \geq 0 \\ \lambda{\cdot}(exp(x) - 1) & ifx < 0 \end{cases}$	

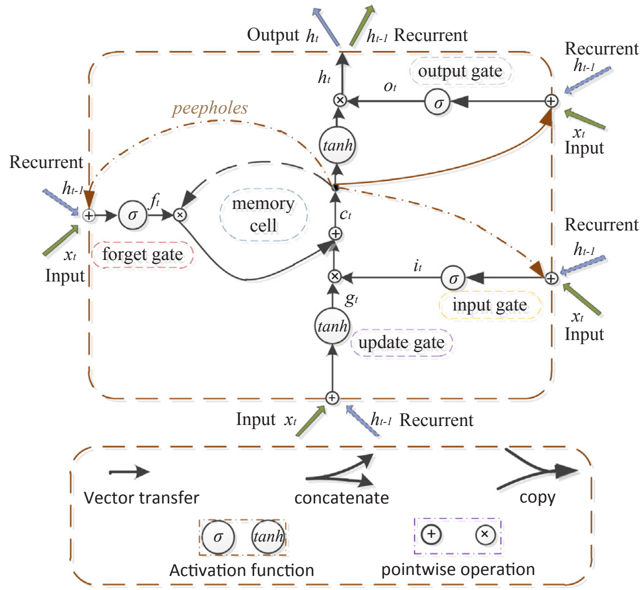


Fig. 4. The specific information dissemination of LSTM.

Table 3
The specific information of this system (1B).

System Specifications	
Array rating	23.4 kW
Panel rating	195 W
PV Technology	mono-Si
Number of panels	4 × 30
Panel type	Trina TSM-195DC01A
Array Area	4 × 38.37 m^2
Type of tracker	DEGERenergie 5000NT, dual axis
Inverter size/type	4 × 6 kW, SMA SMC 6000A
Installation completed	Thu, 8 Jan 2009
Array Tilt/Azimuth	Variable: Dual axis tracking

5.2. Simulation and results

In this section, three experimental simulations were performed to verify the performance of the proposed model and the effect of the input sequence on the prediction accuracy of the models in this paper, including CNN network, LSTM network and the hybrid model (CLSTM). In order to deal with time series problems, 1D CNN was selected in this article. The parameters settings of the proposed network in this paper are shown in Table 4 and were determined by trial and error method. The training and validation epoch in this paper is 100. The CNN model includes convolutional layer, pooling layer, and fully connected layer.

The forecasting results under different input sequences and different models are shown in Table 5. As can be seen from the table, when the input time series length is 0.5 years, the accuracy of the model is the worst, mainly because the input time series length is too short to fully train the model, so the error is high. When the input sequence length is 1 year, the accuracy of the model increases and the accuracy of LSTM model is the highest compared with the other two models. Among them, the MAPE value of LSTM model is 0.103, the MAPE value of CNN model is 0.111, and the MAPE value of CLSTM model is 0.105. Compared with CNN, the indicators of RMSE, MAE and MAPE of LSTM were improved by 10.88%, 3.75% and 7.21% respectively. Compared with CLSTM, RMSE, MAE and MAPE of LSTM were improved by 2.86%, 1.91% and 1.90%, respectively. This is mainly because as the input sequence length increases, the temporal features of the data are gradually extracted by the LSTM model, while the spatial features are relatively weak. CNN and CLSTM cannot extract the spatial features of the data better, so the LSTM model performance is better than the other two models. With the increase of input sequence, the precision of the model increases. When the input sequence length was 3 years, the accuracy of the three models reached the highest, and the accuracy of CLSTM was higher than the other two models. The MAPE value of CLSTM model is 0.022, CNN model is 0.025, and LSTM model is 0.032. Compared with LSTM model, RMSE, MAE and MAPE of CLSTM model increased by 13.82%, 30.39% and 31.25%, respectively. Compared with CNN model, RMSE, MAE and MAPE increased by 6.54%, 10.00% and 12.00%, respectively. Compared with LSTM model, the indexes of CNN RMSE, MAE and MAPE were improved by 7.79%, 22.65% and 21.88%, respectively. The CLSTM model understandably outperformed the other two models. It is understandable that the CLSTM model's

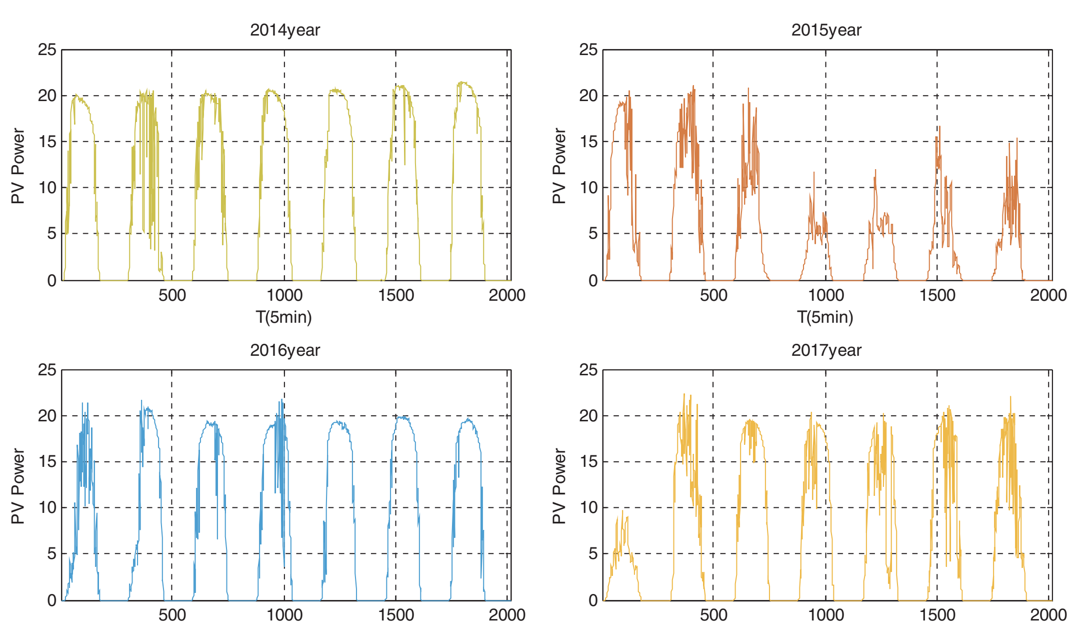


Fig. 5. Partial actual PV power data in 2014–2017 year.

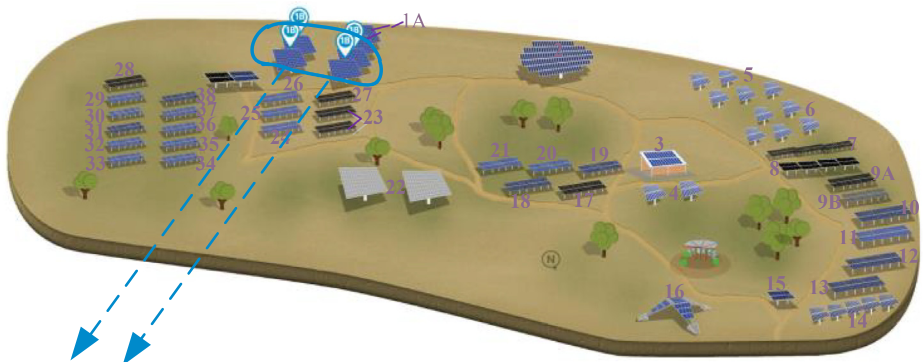


Fig. 6. The map of the system.

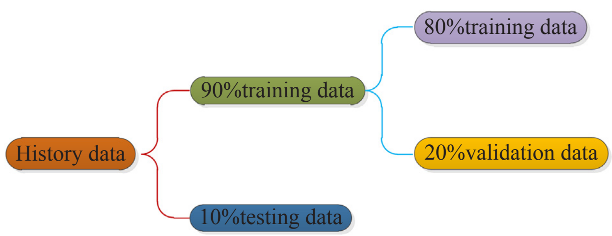


Fig. 7. The data division criteria in this article.

predicted results are better than those of the other two models. This is because for the CLSTM model, it is a hybrid model of CNN and LSTM model, in which the CNN model first extracts local features of data and then extracts the overall timing features of data from the LSTM model, so its prediction results are better than those of the separate CNN and LSTM models. Although the prediction effect of LSTM model is the worst compared with that of the other two models, it can be known from Ref. [43] that its prediction effect is better than that of the traditional shallow neural network. Therefore, the proposed model is not compared with the traditional neural network models in this paper.

However, it can be seen from Table 5 that the accuracy of the model does not increase as the input sequence continues to increase, but decreases as the sequence increases. This may be due to the nature of the time series itself. In terms of time series, not the more data the better the prediction will be. When the data increases to a certain amount, the input data farther from the output is less correlated with the output data, and when the data is increased, the prediction results will be reduced instead of increasing. The RMSE, MAE and MAPE criterion in different input sequence and different models are shown in Fig. 8 to express the performance of the proposed model more intuitively.

Fig. 9 shows the comparison of training and validation error for different input sequence of different models. As can be seen from the figure, when the input sequence is 0.5Y, 1Y and 1.5Y, the errors of each model on the training set and the validation set varies greatly and is in a very unstable state. When the error of the training set is too large, it is easy to cause the under-fitting phenomenon, while when the error of the validation set is too large, it is easy to cause over-fitting phenomenon. With the increase of the input sequence, in the input sequence of 2Y, 2.5Y and 3Y, the effect of each model on the training set and the validation set is very good, but it can be seen from Fig. 9 that the input sequence of 3Y on the testing set shows the best effect. When the input sequence is then increased, the error in the training set and the validation set increases due to the characteristics of the input sequence itself, resulting in an increase in testing error, which is consistent with the results in Fig. 9. Therefore, for different models, the best effect is shown when the input sequence length is 3Y.

Multiple simulation experiments are conducted to prove the stability and robustness of the proposed method. Fig. 10 shows the MAPE, MAE and RMSE results of three models in different input sequence. From Fig. 10, it can be found that in different input sequences and models, whether it is MAPE, MAE or RMSE, the proposed methods all shows high stability and robustness without wide fluctuations and the performance of the CLSTM model is better than that of the other two models. In the input sequence of 3Y shows the best forecasting results. The high stability and robustness may be due to the global optimal solution that deep learning network found. Therefore, it can be concluded that the deep learning forecasting model is highly robust and has

Table 4
The parameters settings in this paper.

Input sequence	Models		
	CNN	LSTM	CLSTM
0.5Y	filters = 512; kernel_size = 2; stride = 2	units = 50;	filters = 200; kernel_size = 2; stride = 2; units = 80;
1Y	filters = 1024; kernel_size = 2; stride = 2	units = 100;	filters = 512; kernel_size = 2; stride = 2; units = 80;
1.5Y	filters = 1228; kernel_size = 2; stride = 2; Dropout = 0.4	units = 150; Dropout = 0.4	filters = 512; kernel_size = 2; stride = 2; units = 80; Dropout = 0.4
2Y	filters = 2048; kernel_size = 3; stride = 2; Dropout = 0.4	units = 100,200; Dropout = 0.4	filters = 1024; kernel_size = 3; stride = 2; units = 80,200; Dropout = 0.4
2.5Y	filters = 3000; kernel_size = 3; stride = 2; Dropout = 0.5	units = 150,500; Dropout = 0.5	filters = 2048; kernel_size = 3; stride = 2; units = 150,300; Dropout = 0.5
3Y	filters = 4096; kernel_size = 3; stride = 2; Dropout = 0.5	units = 80,300,500; Dropout = 0.5	filters = 3000; kernel_size = 3; stride = 2; units = 80,150,300; Dropout = 0.5
3.5Y	filters = 1024,2048; kernel_size = 5; stride = 2; Dropout = 0.6	units = 80,300,600; Dropout = 0.6	filters = 512, 1024; kernel_size = 5; stride = 2; units = 100,300,500; Dropout = 0.6
4Y	filters = 1024,2048,512; kernel_size = 5; stride = 2; Dropout = 0.6	units = 150,500,600; Dropout = 0.6	filters = 512,1024,256; kernel_size = 5; stride = 2; units = 150,200,500; Dropout = 0.6

Table 5
The forecasting results in different input sequence and different models.

Models	LSTM			CNN			CLSTM		
Input sequence	RMSE	MAE	MAPE	RMSE	MAE	MAPE	RMSE	MAE	MAPE
0.5Y	1.244	0.654	0.131	1.128	0.566	0.114	1.161	0.559	0.112
1Y	1.393	0.616	0.103	1.563	0.640	0.111	1.434	0.628	0.105
1.5Y	1.533	0.599	0.101	1.411	0.567	0.095	1.248	0.529	0.095
2Y	1.320	0.457	0.068	0.983	0.452	0.059	0.941	0.397	0.052
2.5Y	0.945	0.389	0.051	0.447	0.231	0.041	0.426	0.198	0.035
3Y	0.398	0.181	0.032	0.367	0.140	0.025	0.343	0.126	0.022
3.5Y	1.150	0.455	0.083	1.136	0.412	0.077	0.991	0.384	0.070
4Y	1.465	0.565	0.089	0.971	0.478	0.083	0.886	0.405	0.080

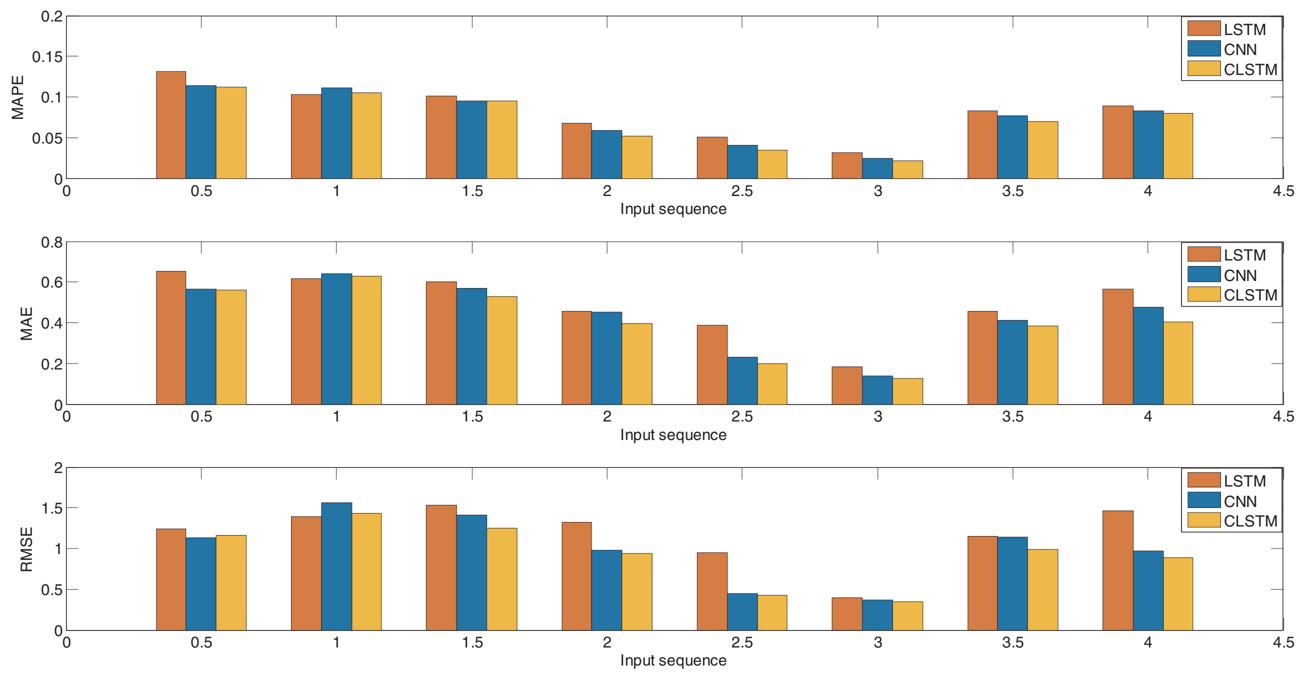


Fig. 8. The RMSE, MAE and MAPE criterion in different input sequence and different models.

a good effect on the PV power prediction model.

Figs. 11–14 shows the comparison of the 5-day PV power prediction results and the actual (the actual value is the measured value, which is considered here as the ‘true value’, i.e. the reference standard value) PV power results for the proposed model in different seasons when the input sequence is 3Y. From the figures, it can be found that, the trend of the prediction results of the three models are roughly the same as the actual results, but the CLSTM model is most similar to the actual prediction result. The CNN model has a better effect than the LSTM model but is lower than the CLSTM model. The values of the selected 5 days of temperature and global irradiance are also shown in Figs. 11–14(b). It can be seen from the figure that temperature and irradiance have a very large effect on PV power generation, especially irradiance, whose trend is very similar to the PV power trend. Considering the powerful ability of the deep learning model to deal with nonlinearity data, in addition to considering the influence of temperature and global irradiance on PV power forecasting, wind speed, wind direction, humidity, diffuse irradiance and other factors that affecting PV power forecasting are also considered.

The running times of different models and different input sequences are shown in Fig. 15. The experimental results were completed in Python3.6 and a personal computer with an 64 bit operating system, Intel (R) Core (7M) i5-8400 CPU@2.8GHZ 2.81GHZ and 8.00 GB of RAM. As can be seen from Fig. 15, as the input sequence increases, the computation time of the model also increases. Although the prediction results of the LSTM model are somewhat worse than the results of CNN and CLSTM, the operation time is the shortest. The CLSTM model has the best prediction results, but its model has the longest running time. When the input sequence is 3Y, the CLSTM runtime is approximately 0.6217 h. In practical applications, when improving the hardware environments or optimizing code, its runtime will be much lower, which is acceptable in practical applications.

6. Conclusion

In this paper, three kinds of photovoltaic power forecasting models (convolutional neural network, long short-term memory neural network and hybrid model) are proposed. By using the time series data of different lengths(in 0.5Y time interval, divided into 8 time periods), a large number of tests and verifications were carried out on them, and the statistical results under three statistical indicators were given. The overall conclusions of this paper are as follows:

- With the increase of available historical time series data, the prediction accuracy of the three models is improved, but when the data length of the time series reaches a certain length, the gain of this precision is slowed down, and even a negative gain appears, which seems that an inflection point occurs at 3Y (for example, for the three models, the time series data length is from 2.5Y to 3Y, the RMSE value is generally increased by 17.00–58.00%, the MAE value

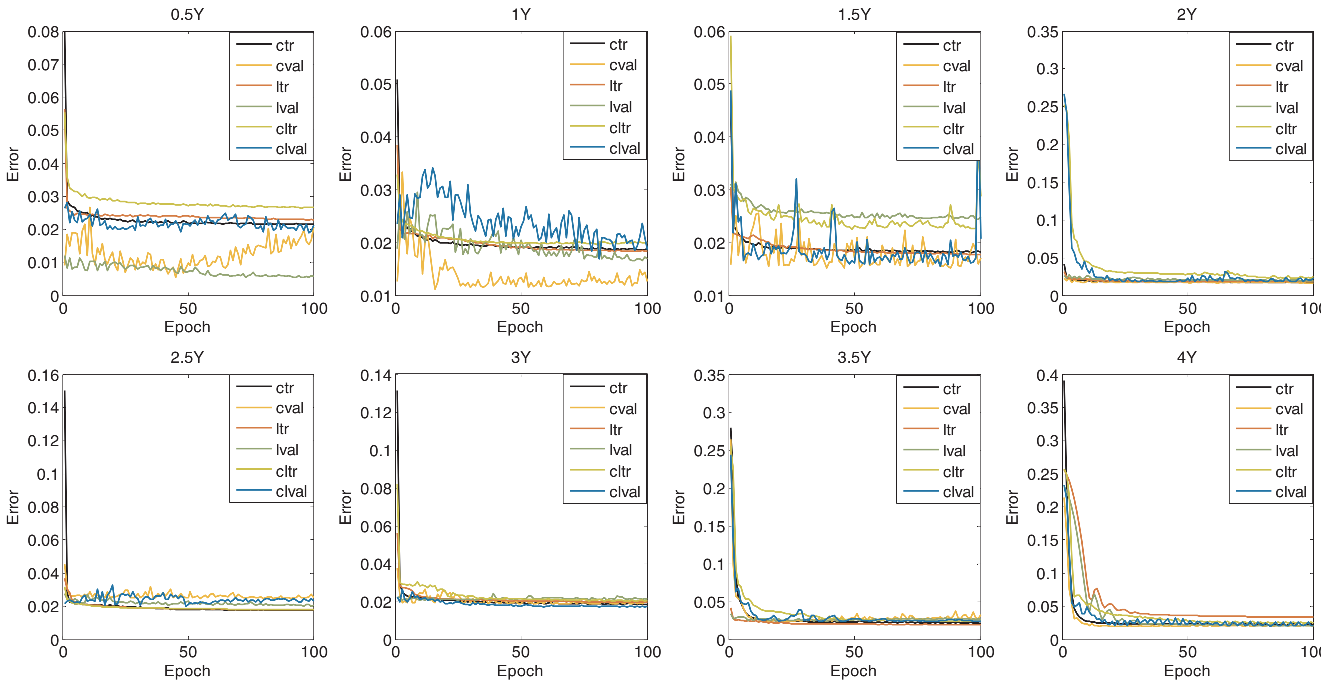


Fig. 9. Comparison of training and validation error for different input sequence of different models.

is generally increased by 36.00–54.00%, and the MAPE value is generally increased by 37.00–40.00%; From 3Yto 3.5Y, the RMSE value is generally increased by −65.00 to −68.00%, the MAE value is generally increased by −60.00 to −68.00%, and the MAPE value is generally increased by −61.00 to −69.00%). Therefore, in this study, it is recommended to select a data length of 3Y for prediction. In view of the universality of the model, it is intended to conduct prediction studies on multiple regions in the future, which will be shown in other studies.

- The statistical results show that all three models perform well and have acceptable accuracy. For a single model, the convolutional neural network model has higher prediction accuracy in 7 time periods, but the long short-term memory neural network model has higher precision in a certain period of time. This seems to indicate that for photovoltaic power prediction, spatial features have a greater advantage over temporal features.
- The statistical results also show that the hybrid model has higher accuracy than the single model (for example, when the time series data length is 2.5Y, 3Y, 3.5Y, etc., compared with the long short-term memory model, the RMSE value of the hybrid model is increased by 54.92%, 13.82%, and 13.83%, respectively; the MAE value of the hybrid model is increased by 49.10%, 30.39%, and

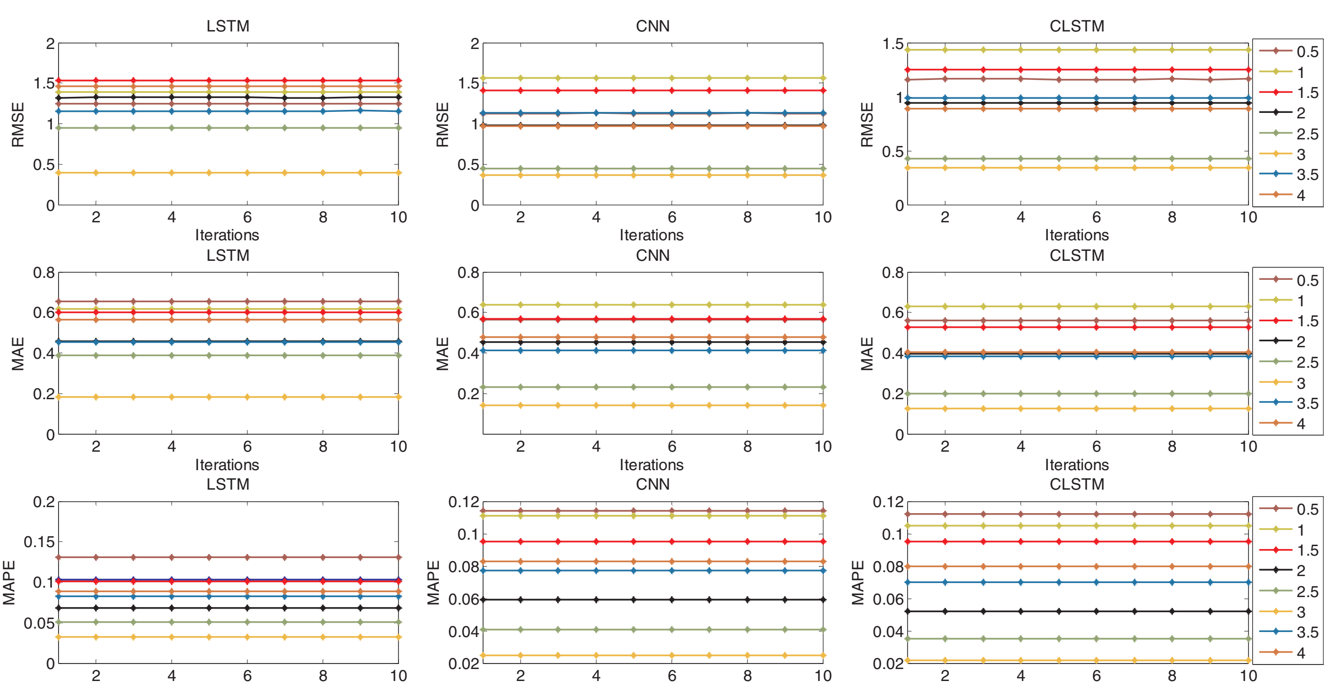


Fig. 10. The MAPE, MAE and RMSE results of three models in different input sequence.

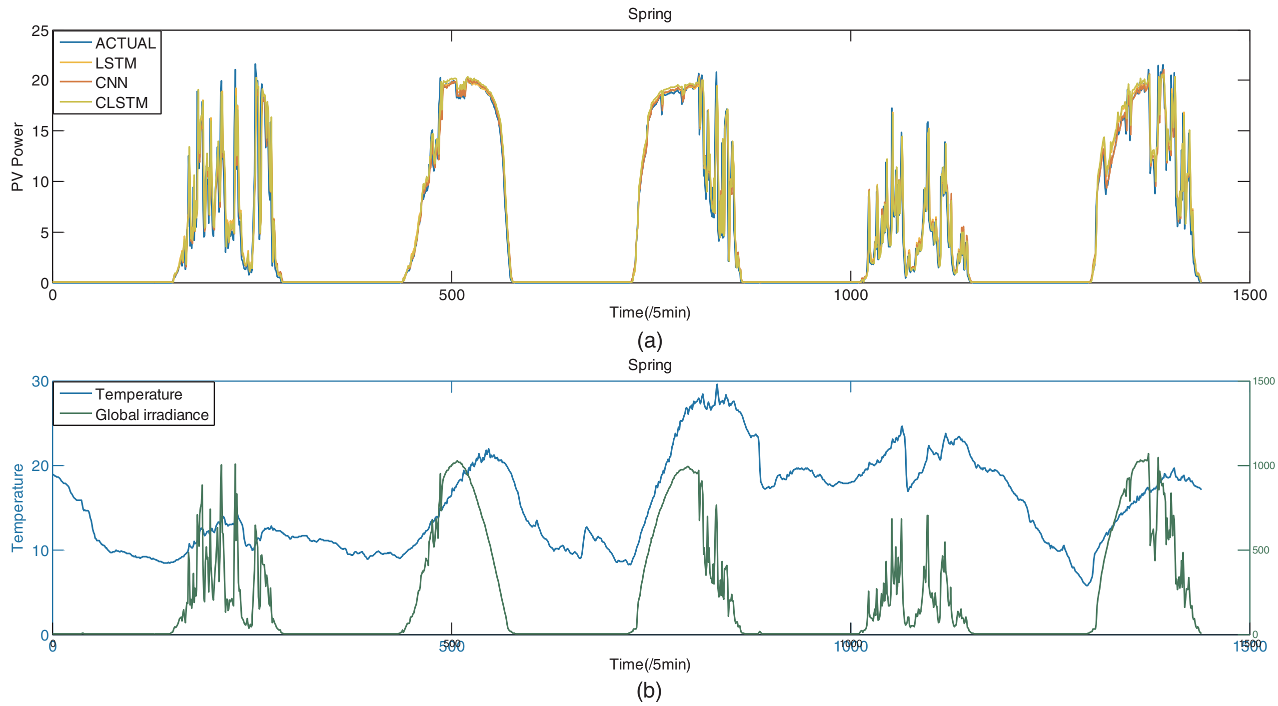


Fig. 11. The forecasting results of the proposed model and actual PV power in spring in the input sequence of 3Y.

15.60%, respectively; the MAPE value of the hybrid model is increased by 31.37%, 31.25%, and 15.66%, respectively. Compared with the convolutional neural network model, the RMSE value of the hybrid model is increased by 4.70%, 6.54%, and 12.76%, respectively; the MAE value of the hybrid model is increased by 14.29%, 10.00%, and 6.80%, respectively; the MAPE value of the hybrid model is increased by 14.63%, 12.00%, and 9.09%, respectively. Median of RMSE, MAE, MAPE for long short-term memory model 13.83%, 30.39%, and 31.25%, Median of RMSE, MAE, MAPE for convolutional neural network model 6.54%, 10.00%, and 12.00%). It is mainly due to the fact that the hybrid model utilizes the respective advantages of the convolutional neural network (which is responsible for extracting the spatial features of the data) and the long short-term memory neural network (which is responsible for extracting the temporal features of the data). This series of results shows that in most cases, hybrid networks work better than a single model. It also shows the value of the hybrid prediction model.

- However, under the condition that the time series data length is 1Y, the prediction accuracy of the long short-term memory neural network models in the three models is the highest in terms of prediction

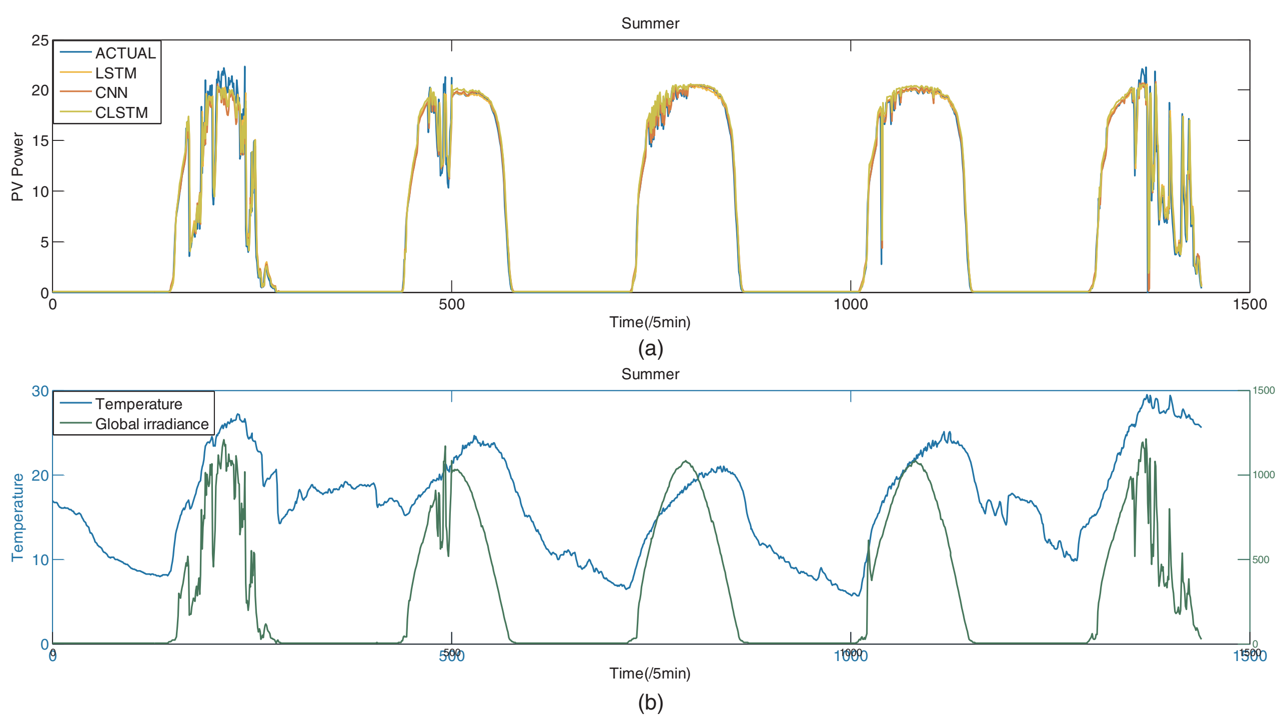


Fig. 12. The forecasting results of the proposed model and actual PV power in summer in the input sequence of 3Y.

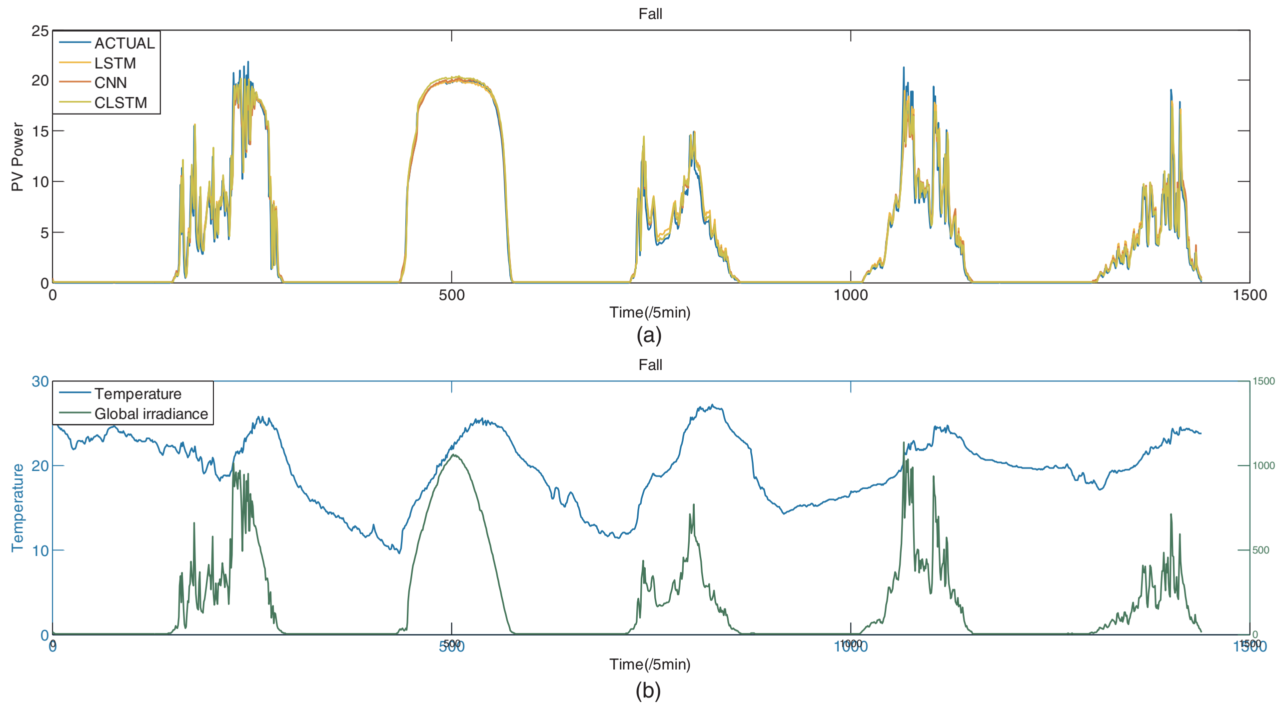


Fig. 13. The forecasting results of the proposed model and actual PV power in fall in the input sequence of 3Y.

accuracy. Mainly because in the process of 0.5Y-1Y, with the increase of data length, the improvement of the prediction effect of long short-term memory networks is more obvious, so the effect is good. In front of it (less than 1Y) and behind, the hybrid network has a better overall effect because it integrates the advantages of two deep networks. Considering the newly built photovoltaic farms with unsound historical data, the long short-term memory neural network model seems to be a good choice when the input data is certain.

- In terms of training time, the long short-term memory neural network has the shortest training time, and the hybrid model has the longest training time. The training time of the hybrid model is approximately 1.5–3.5 times longer than the long short-term memory neural network (varying with time series). However, after training (when used), the model's prediction time is roughly the same. The prediction time of the hybrid model is 2–4 s different from that of the long short-term memory neural network model, while the prediction time of the long short-term memory neural network and the convolutional neural network is basically the same. For a certain length of time series (such as 1.5Y), under certain evaluation criteria

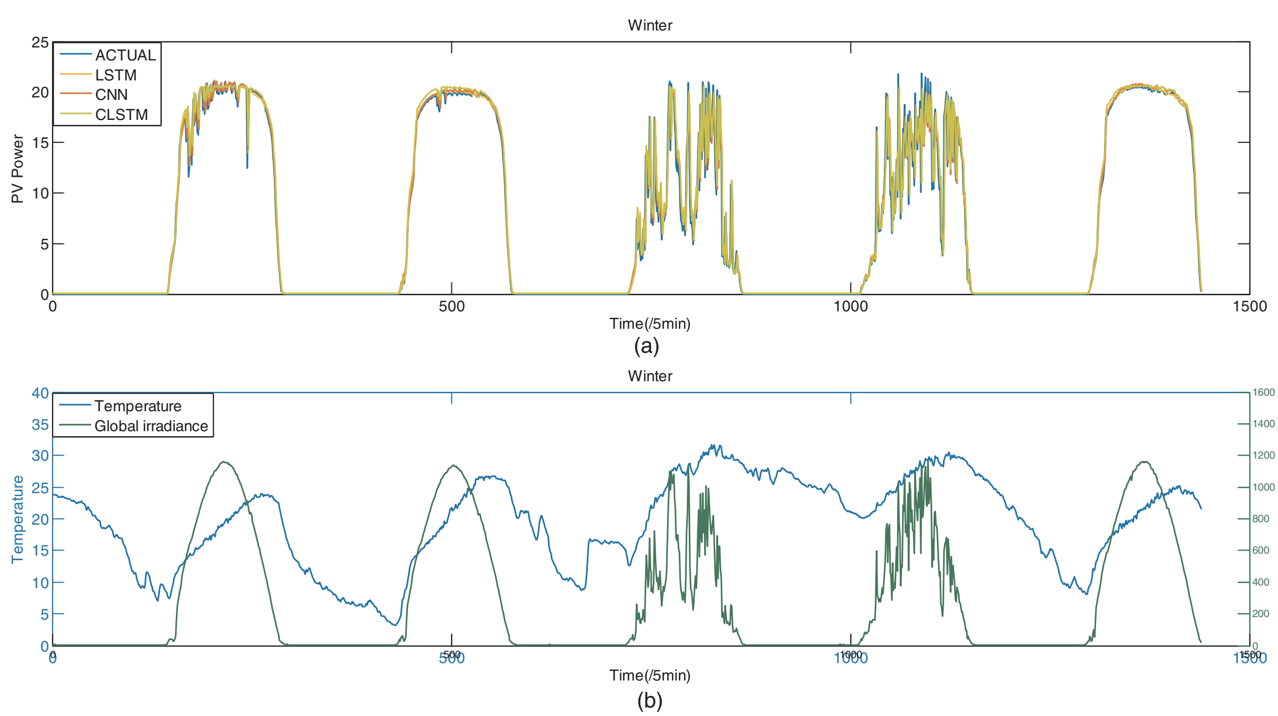


Fig. 14. The forecasting results of the proposed model and actual PV power in winter in the input sequence of 3Y.

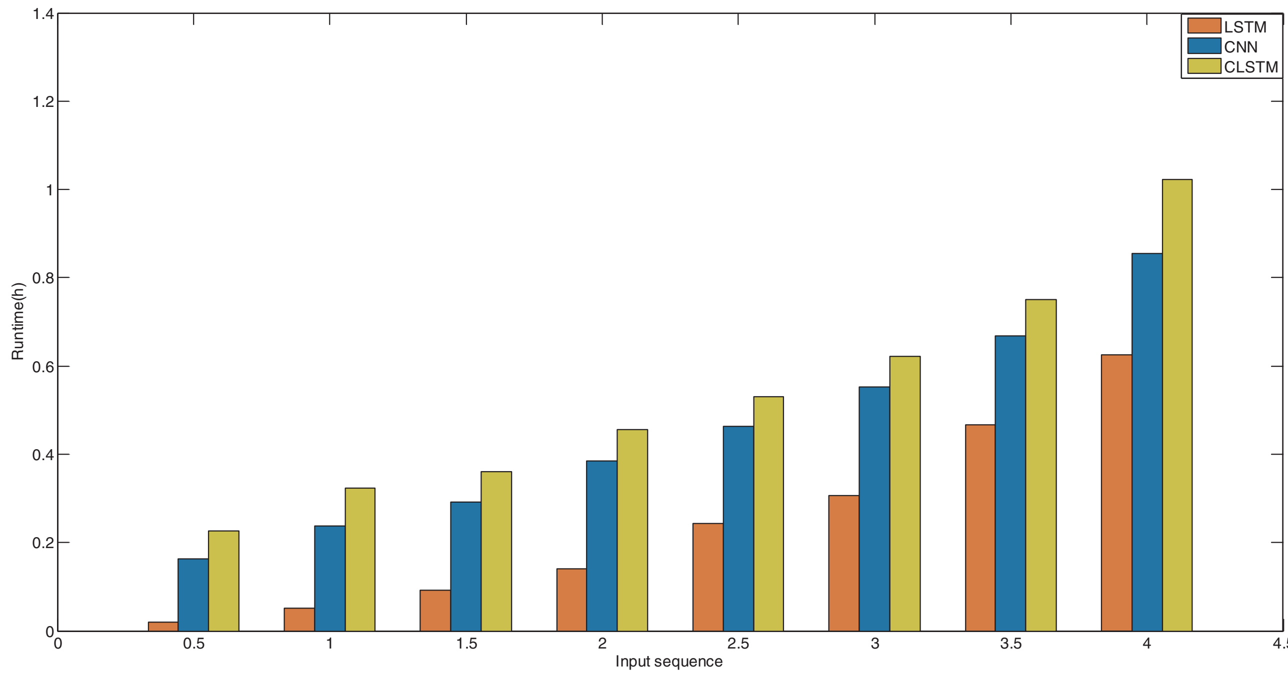


Fig. 15. The runtime of different model and different input sequence.

(such as MAPE), the prediction results of the convolutional neural network are similar to the hybrid model. At the same time, if we consider the advantages of convolutional neural networks in training time, it seems to be a good choice to use the model of convolutional neural network to predict the situation of 1.5Y historical data.

- In summary, in this study, the prediction accuracy of the hybrid model is higher than that of the single model and it is recommended to select 3Y for prediction in the time series data length selection. However, considering the processing time required by the model and the length of historical time series data obtained in the actual situation, convolutional neural networks and long short-term memory neural network models are also good choices under certain circumstances. The research conclusions have important guiding significance for the prediction methods in practical engineering applications and the selection of data lengths for predicting historical time series. Reasonable time series data length can achieve the dual pursuit of high prediction accuracy and low computational cost.

Acknowledgement

This work was supported by Nature Science Foundation of China (51761135013); Nature Science Foundation of Heilongjiang Province and Harbin (E2017017, 2016RQQXJ101); Fundamental Research Funds for the Central Universities (HEUCFP201801).

References

[1] Liu L, Zhao Y, Chang D, et al. Prediction of short-term PV power output and uncertainty analysis. Appl Energy 2018;228:700–11.
[2] van der Meer DW, Shepero M, Svensson A, et al. Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes. Appl Energy 2018;213:195–207.
[3] Changwei Lai, Jinghua Li, Bo Chen, et al. Research review of photovoltaic power generation capacity prediction technology. Trans China Electrotechn Soc 34(6):1201–1217 [Chinese].
[4] Soman SS, Zareipour H, Malik O, et al. A review of wind power and wind speed forecasting methods with different time horizons. North American Power Symposium 2010. IEEE; 2010. p. 1–8.
[5] Dolara A, Leva S, Manzolini G. Comparison of different physical models for PV power output prediction. Sol Energy 2015;119:83–99.
[6] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–88.
[7] Li Y, He Y, Su Y, et al. Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. Appl Energy 2016;180:392–401.
[8] Wei Hou, Jian Xiao, Liyong Niu. Analysis of power generation capacity of photovoltaic power. Electr Eng 2016;17(4):53–8.
[9] Halabi LM, Mekhilef S, Hossain M. Performance evaluation of hybrid adaptive neuro-fuzzy inference system models for predicting monthly global solar radiation. Appl Energy 2018;213:247–61.
[10] Lan H, Yin H, Hong YY, et al. Day-ahead spatio-temporal forecasting of solar irradiation along a navigation route. Appl Energy 2018;211:15–27.
[11] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98.
[12] Voyant C, Notton G, Kalogirou S, et al. Machine learning methods for solar radiation forecasting: a review. Renewable Energy 2017;105:569–82.
[13] Chen C, Duan S, Cai T, et al. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol Energy 2011;85(11):2856–70.
[14] Mellit A, Sağlam S, Kalogirou SA. Artificial neural network-based model for estimating the produced power of a photovoltaic module. Renewable Energy 2013;60:71–8.
[15] Yadav AK, Malik H, Chandel SS. Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76 sites in Northwestern India. Renew Sustain Energy Rev 2015;52:1093–106.
[16] Cervone G, Clemente-Harding L, Alessandrini S, et al. Short-term photovoltaic power forecasting using Artificial Neural Networks and an Analog Ensemble. Renewable Energy 2017;108:274–86.
[17] Almonacid F, Pérez-Higueras PJ, Fernández EF, et al. A methodology based on dynamic artificial neural network for short-term forecasting of the power output of a PV generator. Energy Convers Manage 2014;85:389–98.
[18] Jahani B, Mohammadi B. A comparison between the application of empirical and ANN methods for estimation of daily global solar radiation in Iran. Theor Appl Climatol 2018:1–13.
[19] Gong T, Fan T, Guo J, et al. GPU-based parallel optimization of immune convolutional neural network and embedded system. Eng Appl Artif Intell 2017;62:384–95.
[20] Shi C, Pun CM. Superpixel-based 3D deep neural networks for hyperspectral image classification. Pattern Recogn 2018;74:600–16.
[21] Yao G, Lei T, Zhong J. A review of convolutional-neural-network-based action recognition. Pattern Recogn Lett 2018.
[22] Hinton GE. A practical guide to training restricted Boltzmann machines. Neural networks: tricks of the trade. Berlin, Heidelberg: Springer; 2012. p. 599–619.
[23] Tong C, Li J, Lang C, et al. An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders. J Parallel Distrib Comput 2017.
[24] Rao G, Huang W, Feng Z, et al. LSTM with sentence representations for document-level sentiment classification. Neurocomputing 2018.
[25] Kuremoto T, Kimura S, Kobayashi K, et al. Time series forecasting using a deep belief network with restricted Boltzmann machines. Neurocomputing 2014;137:47–56.
[26] Wan J, Liu J, Ren G, et al. Day-ahead prediction of wind speed with deep feature

learning. Int J Pattern Recognit Artif Intell 2016;30(05):1650011.
[27] Dedinec A, Filiposka S, Dedinec A, et al. Deep belief network based electricity load forecasting: an analysis of Macedonian case. Energy 2016;115:1688–700.
[28] Liu H, Mi X, Li Y. Wind speed forecasting method based on deep learning strategy using empirical wavelet transform, long short term memory neural network and Elman neural network. Energy Convers Manage 2018;156:498–514.
[29] Liu H, Mi X, Li Y. Smart multi-step deep learning model for wind speed forecasting based on variational mode decomposition, singular spectrum analysis, LSTM network and ELM. Energy Convers Manage 2018;159:54–64.
[30] Wang H, Li G, Wang G, et al. Deep learning based ensemble approach for probabilistic wind power forecasting. Appl Energy 2017;188:56–70.
[31] Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018;148:461–8.
[32] Nespoli L, Medici V. An unsupervised method for estimating the global horizontal irradiance from photovoltaic power measurements. Sol Energy 2017;158:701–10.
[33] Schwingshackl C, Petitta M, Wagner JE, et al. Wind effect on PV module temperature: analysis of different techniques for an accurate estimation. Energy Procedia 2013;40:77–86.
[34] Wang H, Yi H, Peng J, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers Manage 2017;153:409–22.
[35] Gu J, Wang Z, Kuen J, et al. Recent advances in convolutional neural networks. Pattern Recogn 2017.
[36] https://github.com/exacity/deeplearningbook-chinese.
[37] http://lamda.nju.edu.cn/weixs/book/CNN_book.pdf.
[38] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. In: Advances in neural information processing systems; 2014. p. 3104–12.
[39] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9(8):1735–80.
[40] Graves A. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850; 2013.
[41] http://dkasolarcentre.com.au/historical-data/download.
[42] http://dkasolarcentre.com.au/locations/alice-springs?source=1B.
[43] Srivastava S, Lessmann S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol Energy 2018;162:232–47.