Article

Combined Prediction of Photovoltaic Power Based on Sparrow Search Algorithm Optimized Convolution Long and Short-Term Memory Hybrid Neural Network

Shun Li , **Jun Yang, Fuzhang Wu, Rui Li *** and **Ghamgeen Izat Rashed**

School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; 2020282070160@whu.edu.cn (S.L.); jyang@whu.edu.cn (J.Y.); wufuzhang@whu.edu.cn (F.W.); ghamgeen@whu.edu.cn (G.I.R.)
* Correspondence: lirui@whu.edu.cn; Tel.: +86-027-17828869575

Abstract: To address the problem of strong uncertainty in the high proportion of new energy output, an improved convolutional long- and short-term memory (CLSTM) hybrid neural network is proposed for PV power combination prediction. Firstly, considering the large impact of weather changes on PV power output, a fluctuation feature identification model is used to classify historical PV power series samples into slow weather change type and severe weather change type. Secondly, taking into account the multimodal characteristics of PV power output, an improved variational modal decomposition technique is used to adaptively determine the number of modal components, *K*, and decompose the two types of samples. Regarding the existence of the low-frequency steady state component and the high-frequency fluctuation component of PV power output, the high-frequency component is used to train the long- and short-term memory (LSTM) model and the low-frequency component is used to train the convolutional neural network (CNN) model. The improved sparrow search algorithm (SSA) is used to optimize the parameters of the LSTM and CNN models during the training process. Finally, the predicted component values of each model are superimposed and reconstructed to obtain PV power prediction values. The actual operation data of a PV plant in northern China were used for comparison and validation, and the experiments showed that the accuracy of the prediction results, based on the improved SSA to optimize the parameters of the CLSTM hybrid neural network for predicting PV output, was significantly better than that of the BP, CNN, LSTM single neural network prediction results, and of the prediction accuracy of the unoptimized CLSTM hybrid neural network. At the same time, compared with the above single neural network and unoptimized hybrid prediction model, the proposed method converged faster and was more adaptable to weather changes.

Keywords: photovoltaic power combination prediction; fluctuation feature recognition model; improved variational modal decomposition; CLSTM hybrid neural network

1. Introduction

The increasing depletion of fossil energy and serious environmental pollution have resulted in an energy crisis, and environmental problems are increasingly prominent [1]. Power systems are the dominant resources to achieve low-carbon goals, and building a new type of power system with a high proportion of renewable energy is the new development direction of the power industry [2]. Photovoltaic power generation, as a promising renewable energy, has the characteristics of cleanliness and sustainability, but it is easily affected by the environment, giving rise to unstable characteristics in its output, such as volatility, randomness, and intermittency [3,4]. As high penetration photovoltaic power generation is connected to the power system, the safety and stability of the operation of the power system becomes greatly challenged. The accurate prediction of photovoltaic output would not only provide powerful data support for its suppression and consumption,

but would also reduce its impact on the power grid, improve its market competitiveness and promote transformation of the power system [5]. Additionally, it is essential for the development of an economical and reliable power dispatch schedule [6,7]. Therefore, it is an important foundation for the safety, stability, and economic operation of the new type of power system that there is high penetration of renewable energy to improve the accuracy of photovoltaic output prediction.

Some research has been carried out on the prediction of photovoltaic output, mainly model and data driven. In terms of model driven research, multiscale analysis, a clustering algorithm and a probability distribution model are used in [8] to realize the mining and modeling of the rapid fluctuation and long-term scale changes of photovoltaic power. A physical model of a photovoltaic power station is established in [9], based on basic information, such as location and placement angle of the photovoltaic system, and combining this with the conversion efficiency formula of photovoltaic modules. However, the photovoltaic power prediction model based on a model-driven method has the disadvantages of modeling with complex and low calculation efficiency. With the integration of a high proportion of photovoltaic power, it is difficult to meet the requirements of power grid regulations on accuracy and rapidity of photovoltaic power predictions. In terms of the data driven model, the emergence and development of artificial intelligence (AI) technology has made photovoltaic power prediction, based on data-driven methods, become a research hotspot; such as support vector machines [10], extreme learning machines [11], Markov chains and other methods. Compared with model-driven methods, the above-mentioned AI methods are more effective.

To further improve prediction accuracy, some researchers have proposed decomposing the mode of the historical photovoltaic power sequence, and then predicting each component, considering that photovoltaic output contains multiple modes. This method is called combined prediction. Existing studies have shown that the accuracy of the combined prediction model is generally superior to a single model. For example, in [12] the authors used empirical mode decomposition to decompose photovoltaic output, but this method has modal aliasing. In [13] the authors introduced variational modal decomposition, and determined the number of modal components by the cumbersome center frequency method; however, it could not adaptively determine the number of modal components. With the continuous development of deep learning and the promotion of applications, deep neural network models, such as CNN and LSTM, have become mainstream methods of photovoltaic power combination prediction. The combined prediction of photovoltaic power is adapted in [14,15], which was based on a convolutional neural network model and long- and short-term memory neural model, respectively. The convolutional long- and short-term memory neural hybrid network model is applied in [16]. Compared with traditional machine learning methods, deep neural network models have obtained more accurate combined prediction results. Although the prediction accuracy of the above methods has been improved, these models, especially deep neural network models, have problems, such as overfitting, gradients exploding or disappearing during training, and slow convergence speed, due to strong fitting ability.

In general, the existing models for photovoltaic output prediction, based on model-driven methods, have problems of relatively low prediction accuracy and complex calculation burdens. The existing data-driven combined prediction models have problems of overfitting and slow convergence, due to defects of the algorithm itself. It is difficult to meet the requirements of high accuracy and rapidity of photovoltaic output forecasting for the operation and control of the renewable power system.

This paper proposes an improved SSA to optimize the parameters of the CLSTM hybrid neural network, which is called the SSA-CLSTM model, and then to establish a photovoltaic power combined prediction model. The photovoltaic output prediction is performed through the CLSTM hybrid neural network to improve prediction accuracy. Meanwhile, the improved SSA is applied to optimize CLSTM parameters to eliminate

over-fitting and improve the convergence speed of the algorithm. The main contributions of this paper are:

(1) The fluctuation feature recognition model is applied to divide historical photovoltaic power sequence samples into slow weather changes and severe weather changes; on this basis, an improved variational modal decomposition technique is used to adaptively determine the number of variational modal components, *K*, then decompose the historical photovoltaic power sequence.

(2) A combined prediction method is proposed, in which high and low frequency components are separately trained and predicted, considering that it is difficult to match the CNN model to the LSTM model. The training and prediction of high-frequency components are implemented by the LSTM model, and the training and prediction of the low-frequency components are carried out with the CNN model.

(3) An improved sparrow search algorithm is proposed to find the best parameters of the CNN and LSTM models to increase convergence speed and improve prediction accuracy.

The rest of the paper is organized as follows. Section 2 introduces the improved variational modal decomposition and fluctuation feature recognition model. Optimizing the entire algorithm flow by improving the initial population method is described in Section 3. The convolutional neural network model and the long and short-term memory network model are introduced in Section 4. The model of the combination prediction model is proposed creatively by using a hybrid neural network in Section 5, which describes in detail the photovoltaic power prediction process using the method proposed in this paper. Case studies and conclusions are presented in Sections 6 and 7, respectively.

2. Improved Variational Mode Decomposition and Fluctuation Feature Recognition Model

2.1. Improved Variational Mode Decomposition

Considering the phenomenon of multimodal aliasing of photovoltaic output, the photovoltaic output modal is first decomposed. The photovoltaic output will show random and periodic fluctuation, and has obvious non-stationary status, due to the influence of the environment and the equipment. Variational mode decomposition (VMD) is a new method, which is different from recursive modal decomposition. It can effectively deal with nonlinear and non-stationary time series signals and has excellent frequency domain decomposition characteristics. When decomposing a signal, the resolution accuracy depends on the number of components parameter, *K*. Existing researchers set the number of components, *K*, in advance, based on their experiences, but determining the number of components based on subjective experiences can easily cause aliasing of modal components. To solve this problem, singular value decomposition technology is introduced. Its clear signal-to-noise resolution capability can automatically find the number of modal components according to the best effective order of singular values, and then performs the variational modal according to the determined number of modal components decomposed [17].

2.2. Fluctuation Feature Recognition Model

To accurately determine whether the weather has abrupt changes, the fluctuation feature recognition model in [18] is introduced.

The following daily photovoltaic power sequence fluctuation characteristic parameters are defined for improving the accuracy of fluctuation characteristics recognition.

(1) Peak value of daily photovoltaic power sequence fluctuation M_m:

$$M_\mathrm{m} = \max(\left|\overline{M}'_{r+1} - \overline{M}'_r\right|) \quad r = 1, 2, \cdots, n-1 \tag{1}$$

where $\overline{M}'_r$ indicates the value of the daily photovoltaic power extreme point sequence at time *r*, after normalization; *r* indicates the serial number of the normalized daily photovoltaic power extreme value point sequence; *n* indicates the number of extreme points of the daily photovoltaic power sequence.

(2) Daily photovoltaic power sequence fluctuation frequency f:

$$f = \frac{n}{96} \tag{2}$$

(3) Daily fluctuation rate of photovoltaic power sequence η_m:

$$l_r = \begin{cases} t_r & r = 1 \\ t_r - t_{r-1} & r = 2, 3, \cdots, n \end{cases} \tag{3}$$

$$\eta_m = \max l_r \quad 1, 2, \cdots, n \tag{4}$$

where l_r indicates the time interval between adjacent extreme points of the normalized daily photovoltaic power sequence; t_r indicates the time corresponding to the extreme point of the normalized daily photovoltaic power sequence.

According to the defined fluctuation characteristic parameters of daily photovoltaic power sequence, the characteristic vector of daily photovoltaic power sequence fluctuation is recorded as $W = [M_m, f, \eta_m]$. The quartile method is used to determine the threshold of characteristic parameters. The calculation steps are as follows:

(1) Calculate the second quartile $Q_{2,K,i}$ of the fluctuation characteristic parameters of each day's photovoltaic power sequence:

$$Q_{2,K,i} = \begin{cases} \omega_{(H+1)/2,K,i}, H = 2h+1, h = 0, 1, 2 \cdots \\ \frac{\omega_{H/2,K,i} + \omega_{(H+2)/2,K,i}}{2}, H = 2h, h = 1, 2, \cdots \end{cases} \tag{5}$$

where $\omega_{(H+1)/2,K,i}$ represents the $(H+1)/2$-th value in the k-th day photovoltaic power fluctuation characteristic parameter sequence $\omega_{K,i}$ under the i-th weather type, $K = 1, 2, 3$, $i = 1, 2$.

Calculate the first quartiles $Q_{1,K,i}$ and the third quartiles $Q_{3,K,i}$ of the characteristic parameters of photovoltaic power fluctuations on each day.

When the number of sample days is $H = 2h(h = 1, 2, \cdots)$, dividing the daily photovoltaic power fluctuation feature vector into two parts from $Q_{2,K,i}$, the two parts do not contain $Q_{2,k,i}$, and then calculate the median $Q'_{2,K,i}$ and $Q''_{2,K,i}$ of the two parts respectively, and the former is smaller than the latter, then $Q_{1,K,i} = Q'_{2,K,i}$, $Q_{3,K,i} = Q''_{2,K,i}$.

When the number of sample days is $H = 4h + 3(h = 0, 1, \cdots)$, then:

$$\begin{cases} Q_{1,K,i} = 0.75\omega_{h+1,K,i} + 0.25\omega_{H+2,K,i} \\ Q_{3,K,i} = 0.25\omega_{3h+2,K,i} + 0.75\omega_{3h+3,K,i} \end{cases} \tag{6}$$

When the number of sample days is $H = 4h + 1(h = 0, 1, \cdots)$, then:

$$\begin{cases} Q_{1,K,i} = 0.25\omega_{h,K,i} + 0.75\omega_{H+1,K,i} \\ Q_{3,K,i} = 0.75\omega_{3h+1,K,i} + 0.25\omega_{3h+2,K,i} \end{cases} \tag{7}$$

Calculate the interquartile moment, and determine the threshold of each parameter of the daily photovoltaic power fluctuation feature vector.

The expression of the interquartile moments for calculating the fluctuation characteristic parameters of the two weather types is:

$$I_{QR,K,i} = Q_{3,K,i} - Q_{1,K,i} \tag{8}$$

The expression for determining the fluctuation characteristic parameter sequence of the two weather types is:

$$[F_{1,K,i}, F_{u,K,i}] = \left[Q_{1,K,i} - 1.5I_{QR,K,i}, Q_{3,K,i} + 1.5I_{QR,K,i}\right] \tag{9}$$

where $F_{1,K,i}$ represents the lower threshold of the fluctuation characteristic parameter of the k-th daily power sequence under the i-th weather type; $F_{u,K,i}$ represents the higher threshold of the fluctuation characteristic parameter of the k-th daily power sequence under the i-th weather type, the sample data outside the interval $[F_{1,K,i}, F_{u,K,i}]$ are all outliers.

The improved variational modal decomposition technique is used to decompose the historical photovoltaic power sequence and several modal components are obtained. Among them, high-frequency components are considered to be the sequence with more complex fluctuation characteristics. However, it is difficult to directly distinguish high and low frequency components in practice.

3. Improved Sparrow Search Algorithm Theory

3.1. Improved Tent Chaotic Map

Many swarm intelligence optimization algorithms, such as PS0, determine the initial position of the individual through random initialization. Although the randomness of the initial position of the individual can be guaranteed, it may make the initial position of the individual farther from the optimal position, resulting in low solution accuracy and slow convergence. The chaotic mapping sequence has ergodicity, randomness, and sensitivity to initial values, which can effectively make up for the shortcomings of the random initialization method. The expression of Tent chaotic map is as follows:

$$z_{k+1} = \begin{cases} 2z_k & 0 \leq z_k \leq 0.5 \\ 2(1 - z_k), & 0.5 \leq z_k \leq 1 \end{cases} \quad k = 0, 1, 2, \cdots \tag{10}$$

where k indicates the number of mappings; z_k represents the k-th mapping value.

It can be seen from the (10) that the chaotic map belongs to the distribution of [0, 1], but existing studies have shown that its actual distribution is mainly concentrated between [0.2, 0.8], so the Tent chaotic map is improved, and the improved function expression is as follows:

$$z_{k+1} = \begin{cases} 2(z_k + 0.1 \times rand(0,1)), & 0 \leq z_k \leq 0.5 \\ 2(1 - z_k - 0.1 \times rand(0,1)), & 0.5 \leq z_k \leq 1 \\ k = 0, 1, 2 \cdots \end{cases} \tag{11}$$

In (11), when $z_{k+1} > 1$, its return value is 1; when $z_{k+1} < 0$, its return value is 0.

The ergodicity of the improved Tent chaotic map between [0, 1] is significantly improved, and the distribution is more uniform. Therefore, this paper uses the improved Tent chaotic map to initialize the population.

The inverse mapping of chaotic variables into population solution space variables is as follows:

$$x_k = l_k + (u_k - l_k) \cdot z_k, k = 0, 1, 2, \cdots \tag{12}$$

where l_k and u_k represent the minimum and maximum values of the optimized variable interval, respectively.

Population initialization process:

Step1: Set the number of optimization variables to N, and randomly assign N initial values z_0 of z_k in (10);

Step2: Use (11) to generate chaotic variables $\{z_{ik}, i = 1, 2, ..., n\}$;

Step3: Use (12) to inversely map the chaotic variable z_{ik} to the population solution space to complete the population initialization.

3.2. Sparrow Search Algorithm

SSA is mainly inspired by the foraging behavior and antipredation behavior of sparrows. It is a novel algorithm, with the advantages of strong optimization ability and fast convergence speed. In SSA, all individuals are divided into discoverers, joiners and guarders. The discoverer mainly provides foraging directions for the population, the joiner follows the discoverer for food, and the guarder is responsible for supervising the foraging

area. If danger is detected, the entire population will be notified to flee immediately. During the foraging process, the three individuals constantly update their current positions to find the optimal position. The proportion of guarders in the population is 15~20%, and the discoverers and joiners are dynamically changing. When an individual becomes a discoverer, another individual must become a follower.

Suppose there are N sparrows in the population, the population composed of all individuals can be expressed as $X = [x_1, x_2, \ldots, x_N]$, and the fitness function corresponding to each individual is $F = [f(x_1), f(x_2), \ldots, f(x_N)]^T$, where the discoverer location update rule is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \exp(\frac{-i}{\alpha \times iter_{\max}}), R_2 < ST \\ x_{i,j}^t + Q \times L, R_2 \geq ST \end{cases} \tag{13}$$

where t represents the current iteration number; $x_{i,j}^t$ represents the position of the i-th sparrow in the j dimension in the t-th generation; $\alpha \in (0, 1]$; $iter_{\max}$ is the maximum number of iterations, R_2 is the alarm value, ST is the safety threshold, Q is a random number that obeys a normal distribution; L is a $1 \times d$ matrix, and d represents an individual dimension.

The follower's location update rules are as follows:

$$x_{i,j}^{t+1} = \begin{cases} Q \times \exp(\frac{x_w^t - x_{i,j}^t}{i}), i > \frac{n}{2} \\ x_P^{t+1} + \left| x_{i,j}^t - x_p^{t+1} \right| \times A^+ \times L, i \leq \frac{n}{2} \end{cases} \tag{14}$$

where x_w^t represents the position of the individual with the worst fitness in the t-th generation; X_p^{t+1} represents the position of the individual with the best fitness in the $t + 1$ generation; A is a $1 \times d$ matrix, where each element is randomly set to ± 1, $A^+ = A^T(AA^T)^{-1}$.

The location update rules of the guarder are as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_b^t + \beta \left| x_{i,j}^t - x_b^t \right|, f_i \neq f_g \\ x_b^t + k(\frac{x_{i,j}^t - x_b^t}{|f_i - f_w| + \varepsilon}), f_i = f_g \end{cases} \tag{15}$$

where x_b^t is the global optimal position in the t-th iteration; β controls the step size and obeys the normal distribution of (0,1); $k \in [-1, 1]$; ε is set as a constant to avoid the denominator being zero; f_i is the current individual fitness value, fg and f_w are the current global optimal and worst individual fitness values respectively.

3.3. Improved Sparrow Search Algorithm Optimization Process

In order to improve the prediction accuracy of the CLSTM hybrid network model proposed in the article, an improved SSA is used to optimize the CLSTM hybrid network model. The optimization process is shown in Figure 1.
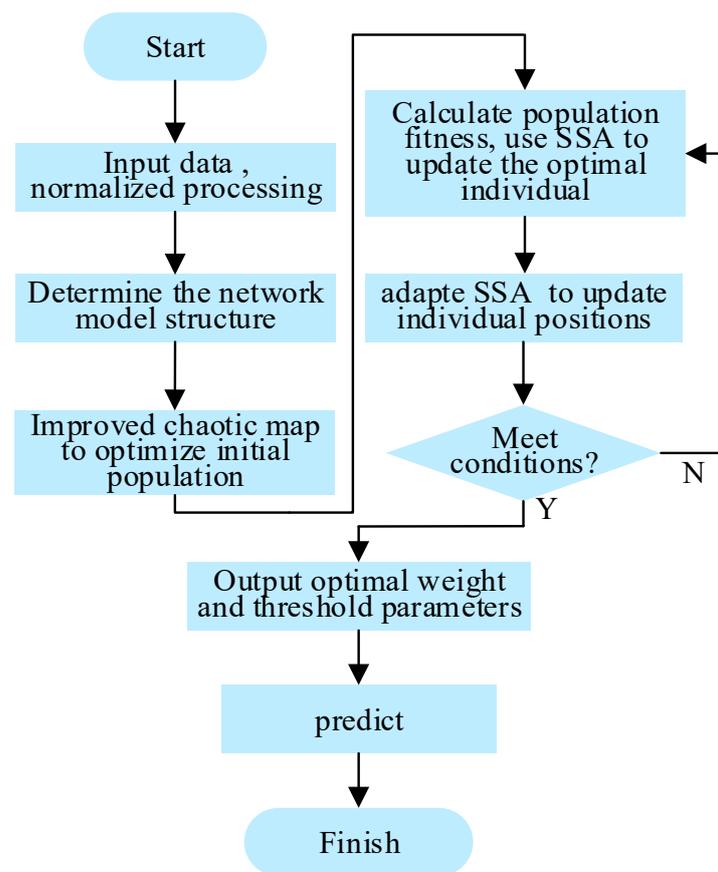
Figure 1. Flow chart of improved sparrow search optimization algorithm.

4. CNN/LSTM Hybrid Network Model

The accuracy of CNN and LSTM in photovoltaic power prediction is acceptable, but when dealing with more complex problems, the accuracy of CNN is relatively slow compared with LSTM. Considering that the low-frequency components have more regular fluctuation forms and the model parameters are relatively less, choosing CNN to build a prediction model can not only meet prediction accuracy but also reduce computational complexity of the model; and the prediction of the high-frequency photovoltaic components can be considered a more complicated problem. At this time, CNN cannot meet its prediction accuracy, and more accurate prediction results can be obtained by using the LSTM network. Therefore, the combined prediction model proposed in this paper is that the low-frequency component prediction model is established by the CNN, and the high-frequency fluctuation component prediction model is established by the LSTM.

4.1. Long and Short-Term Memory Neural Network

LSTM is a special neural network model improved on the basis of RNN. The difference is that forget gates, input gates, update gates, and output gates are added to each hidden layer in RNN. In this way, the current information can be compared with historical information, and learning can be carried out through the mechanism of choosing, forgetting and self-decision, which can alleviate the problem of gradient explosion or disappearance in RNN training. The unit structure diagram is shown in Figure 2.
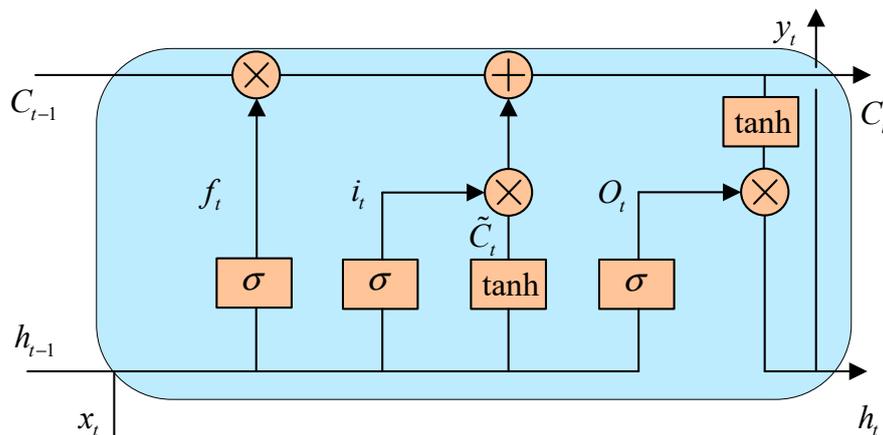
Figure 2. LSTM network model structure diagram.

The model has three inputs, including C_{t-1}, h_{t-1} and x_t, which represent the long-term memory information of the previous moment, the short-term memory information of the previous moment, and the current input, respectively. There are three gates inside the model to control whether the information is discarded or not: input gate, output gate, and forget gate. The update formula is as follows:

$$
\begin{cases}
f_t = \sigma(w_f \cdot [h_{t-1}, x_t] + b_f) \\
i_t = \sigma(w_i \cdot [h_{t-1}, x_t] + b_i) \\
O_t = \sigma(w_o \cdot [h_{t-1}, x_t] + b_o) \\
\widetilde{C}_t = \tanh(w_c \cdot [h_{t-1}, x_t] + b_c) \\
C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \\
h_t = O_t * \tanh(C_t) \\
y_t = w_y h_t + b_y
\end{cases}
\tag{16}
$$

where w and b respectively represent the weight matrix and bias vector of the control gate; σ is the activation function; y_t is the final output result; * is the Hadamard product.

4.2. Convolutional Neural Network

Convolutional neural networks have many applications in image recognition, classification and prediction. The model structure is mainly composed of a convolutional layer and a pooling layer. It uses the characteristics of local connection and weight sharing to speed up training and improve generalization performance. Its typical structure diagram is shown as in Figure 3.
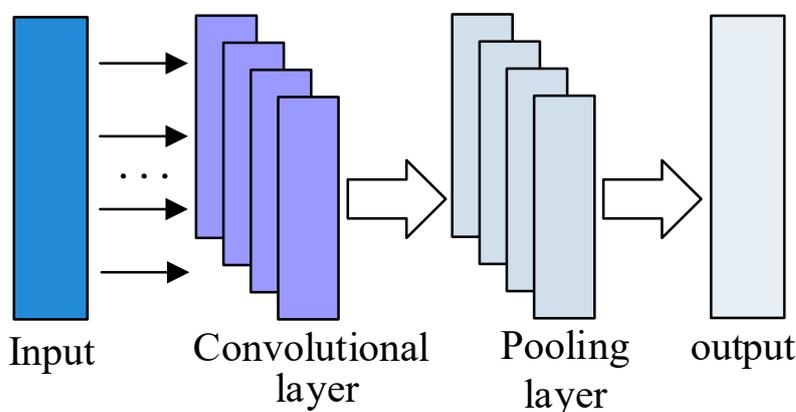

Figure 3. Typical structure diagram of convolutional neural network.

The convolutional layer is composed of multiple feature surfaces, and each feature surface is composed of multiple neural nodes (cells). Each neuron is associated with the corresponding local feature area of the previous layer through the convolution kernel. The feature is extracted through the convolution operation of the convolution kernel. Following the convolutional layer is the pooling layer. Similar to the convolutional layer, the pooling layer is also composed of multiple feature surfaces, each of which corresponds to the feature surface of the previous layer, so the number of feature surfaces is the same as that of the convolutional layer. Meanwhile, the function of the pooling layer is to extract data features twice, and the purpose is to reduce the dimensionality of the data. The commonly used pooling methods are the average pooling method and the maximum pooling method.

CNN is a neural network specially used to process data with a known grid-like topology. In the application scenario of this paper, one-dimensional time series data needs to be processed. One-dimensional time series data can be regarded as a one-dimensional grid sampled at a certain time interval. Therefore, this paper uses a one-dimensional convolutional neural network. The one-dimensional convolution calculation expression is:

$$x_k^l = f(\sum_{i=1}^{N} x_i^{l-1} * w_{ik}^l + b_k^l) \tag{17}$$

where x_k^l represents the k-th convolution of the layer i; f is the activation function; N represents the number of convolutional mappings of the input data; * is the convolution operation; w_{ik}^l is the weight of the i-th operation for the k-th convolution kernel of the layer i; b_k^l is the offset of the k-th convolution kernel corresponding to the layer i.

This paper adopts the maximum pooling method, the expression is $\hat{x}_k^l = \max(x_k^l : x_{k+r-1}^l)$, which means to take the maximum value from vector x_k^l to vector x_{k+r-1}^l. For the sequence x, the maximum pooling operation is repeated for each continuous vector whose window is r, and the maximum feature sequence can be obtained.

5. The Prediction Process of PV Power Based on SSA—CLSTM

The photovoltaic power prediction process proposed in this paper, as shown in Figure 4, is mainly divided into the following steps:

(1) Using the fluctuation feature recognition model, the historical photovoltaic power sequence samples are divided into two types, slow weather changes and severe weather changes.

(2) Using singular value decomposition technology, historical photovoltaic power sequence is decomposed to determine the optimal component parameter, *K*.

(3) Carrying out variational modal decomposition for the two types of weather respectively, the modal components are obtained. Assume that there are *K* modal components corresponding to the slow weather type.

(4) The *K* modal components obtained in 3) can be divided into high-frequency components and low-frequency components through direct observation through experiences.

(5) A CNN prediction model for low-frequency components and a LSTM prediction model for high-frequency components are built. Specifically, component M_1 to component M_N are used as the training output of CNN_1 to CNN_N models, and component M_{N-1} to component M_K are used as the training output of $LSTM_{N-1}$ to $LSTM_K$ models. All models use the same historical environment data as the corresponding training input. In the process of training the model, the improved sparrow search algorithm is used to optimize the model parameters, so as to establish the corresponding prediction model.

(6) Inputting the environmental data of the current period into the trained prediction model, and superimposing and reconstructing the prediction results of the high and low frequency components the predicted value of photovoltaic power is obtained.
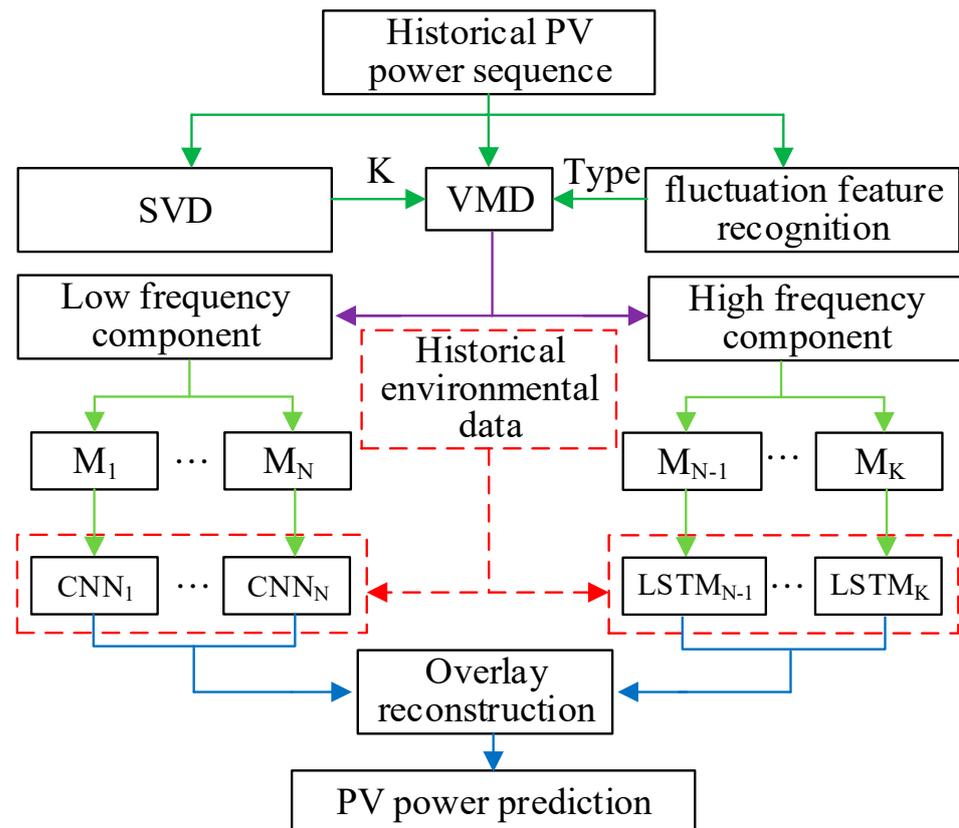
Figure 4. PV power prediction process.

6. Case Studies and Discussion

The following computer configuration is applied in case studies. The processor is Intel(R) Core (TM) i5-4200UCPU@1.60 GHz 2.3 GHz, 64-bit operating system, X64-based processor, and the simulation software is MATLABR2019b. The predictions of the improved CLSTM model proposed in the paper, the improved SSA optimization algorithm model, the two single neural network models of CNN and LSTM, the unoptimized CLSTM model are compared.

Due to the large scale and dimension of the data set provided in the studies, the training time of each model is long, so the performance indicators of various model prediction errors are mainly considered.

6.1. Data Set

The data set in case studies comes from competition data from a Chinese company. The size of the data set is 3671×19. The first 18 dimensions are the environmental factors that affect photovoltaic output and the operating parameters of photovoltaic equipment, and the last dimension is the corresponding historical photovoltaic power value. In order to verify the prediction performance of the model proposed in the paper, 3500 samples of the experimental data set, prepared in the previous stage, are used as the training set, and 171 samples are used as the test set.

6.2. Error Measure

When evaluating the accuracy of the prediction model, the average square error (MSE), average absolute error (MAE), average square root error (RMSE) and average absolute error rate (MAPE) are generally selected as the error evaluation indicators. The calculation expressions are as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \tag{18}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i| \tag{19}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2} \tag{20}$$

$$MAPE = \frac{100\%}{N}\sum_{i=1}^{N}\left|\frac{x_i - \hat{x}_i}{\hat{x}_i}\right| \tag{21}$$

In (18)–(21), indicates the predicted value and actual value of photovoltaic power after pretreatment; N is the number of the data in the test set.

6.3. Improved Variational Modal Decomposition Results

The 3500 historical photovoltaic output power sequence samples are selected from the test set, as shown in Figure 5. It can be concluded from Figure 5 that the photovoltaic output power sequence has obvious characteristics of volatility, randomness, and non-stationarity. If the fluctuation feature recognition model is not used to distinguish the types, and all historical photovoltaic power sequences are directly decomposed into samples, the decomposed modal component diagrams are shown in Figure 6. Obviously, it can be observed that there are many modal components and complex fluctuation forms, so the training is difficult, which will eventually affect prediction accuracy.
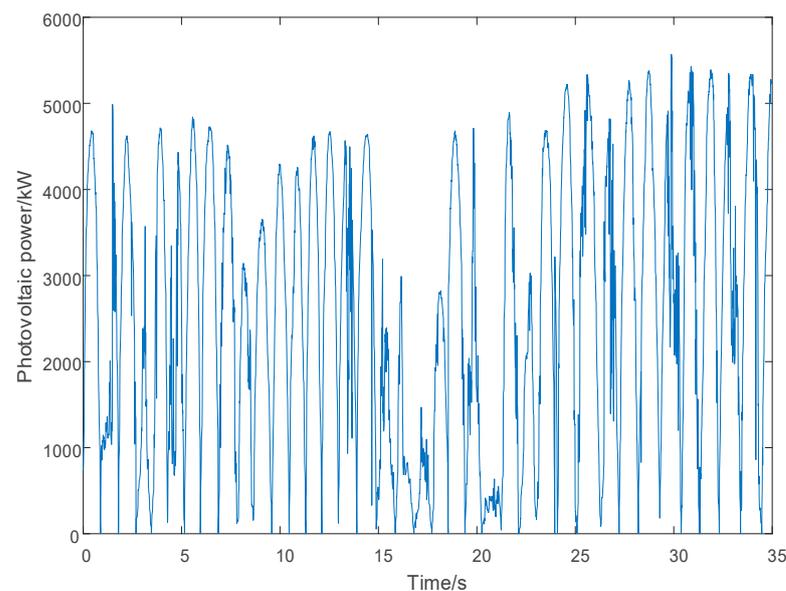

Figure 5. PV historical power time series.

The fluctuation feature recognition model is used to divide the selected samples into two types, slow weather changes and severe weather changes; then, the improved variational modal decomposition technology is applied to decompose the two types of data. The samples of slow weather changes and severe weather changes are decomposed separately, as shown in Figures 7 and 8. Compared with Figure 6, it can be seen that the modal components in Figures 7 and 8 are less and the fluctuation form is regular, which is convenient for subsequent training and prediction.
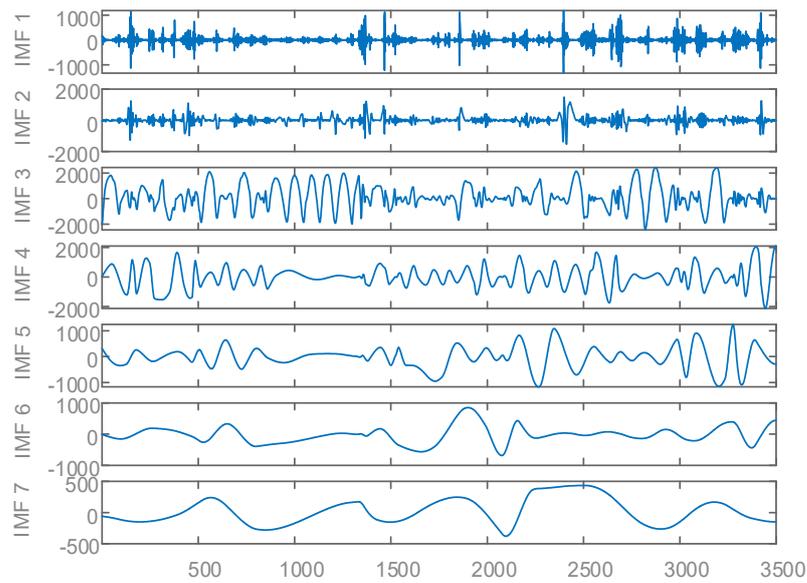
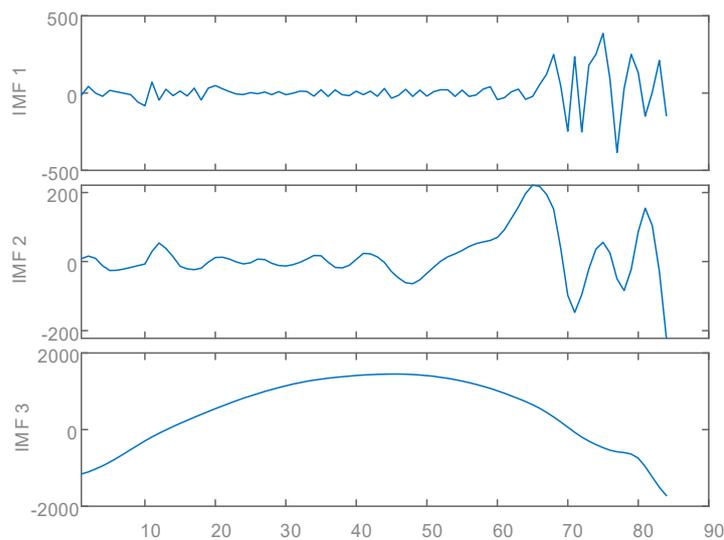**Figure 6.** Improved VMD decomposition results of all samples.

Figure 7. Exploded view of samples with slow weather changes.

In addition, improved variational modal decomposition technology can adaptively decompose the sample sequence, which can avoid the cumbersome center frequency method used in ref. [13] to find the appropriate number of modes.

6.4. Comparative Studies

Through the processing in Section 6.3, it can be seen from Figure 7 that component one is a more complex signal, components two and three are stationary signals; so, component one is used to train the LSTM model, and components two and three are used to train the CNN model. It can be obtained from Figure 8 that components one and two are more complex signals, components three, four, five, and six are stationary signals; so, components one and two are used to train the LSTM model, and components three, four, five, and six are used to train the CNN model.
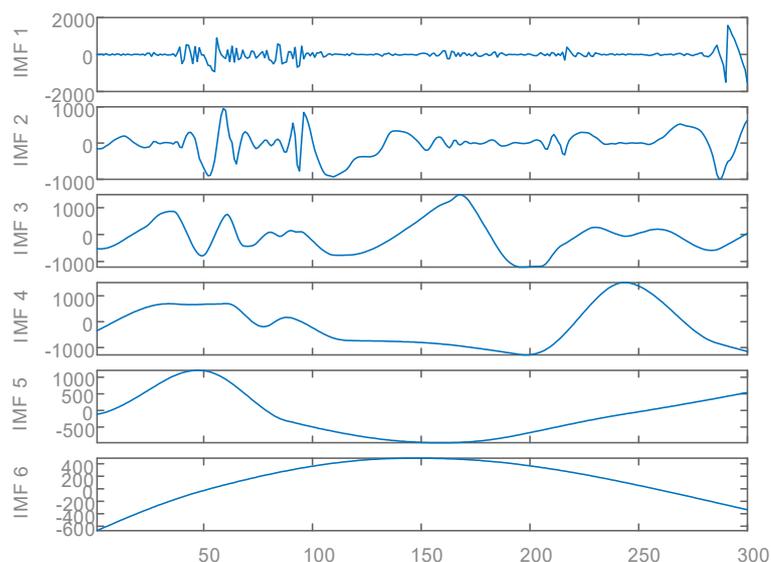
Figure 8. Types of samples with severe weather changes exploded view.

In order to verify the superiority of the models used in the paper, five types of test sets, namely, BP model, CNN model, LSTM model, unoptimized CLSTM model, and optimized CLSTM model are used to test the weather changes slowly and severely. The comparison of the prediction result curves of different models under the two weather types is shown in Figure 9. The error comparison of the different models under the two weather types is shown in Table 1.

In terms of prediction accuracy, from the comparison of the predicted power curve of various network models in the two categories (a) and (b) with the real power curve, the prediction effect of the BP network model is the worst, the prediction accuracy of the CNN and LSTM network are significantly better than that of the BP network model, the LSTM network model is better than the CNN network model; the unoptimized CLSTM hybrid network model is better than the LSTM network model; and the method proposed by this paper obtains the best effect, especially when the weather changes drastically.

Through further specific analysis of the prediction error indicators, it is found that, regardless of the weather type, the BP network model has the worst prediction effect under various error indicators. This is mainly due to the fact that the BP network model's fitting ability is poor and is especially evident under the MSE indicator, on the basis that the indicator error is already very small. The SSA-CLSTM hybrid network model of the two weather types reduces the error by 50% and 82% respectively compared with the single network model. The errors are reduced by 5% and 9% to the unoptimized CLSTM hybrid network model. This further shows that the SSA-CLSTM model proposed in this paper has not only higher photovoltaic prediction accuracy, but also has outstanding adaptability to the impact of weather.

Compared with a single network model, the prediction accuracy of the hybrid network model is significantly improved, which fully shows that the hybrid network model has the same ability as LSTM when dealing with sequences with obvious rules; and when dealing with complex tasks, the learning ability of LSTM is better. Compared with the unoptimized hybrid network model, it reflects the powerful ability of SSA training parameters, which not only avoid over-fitting of the model training, but also prevent the gradient disappearing or exploding during the training, thereby improving prediction accuracy. Therefore, the SSA-CLSTM hybrid network model performs best in the analysis of various error indicators.
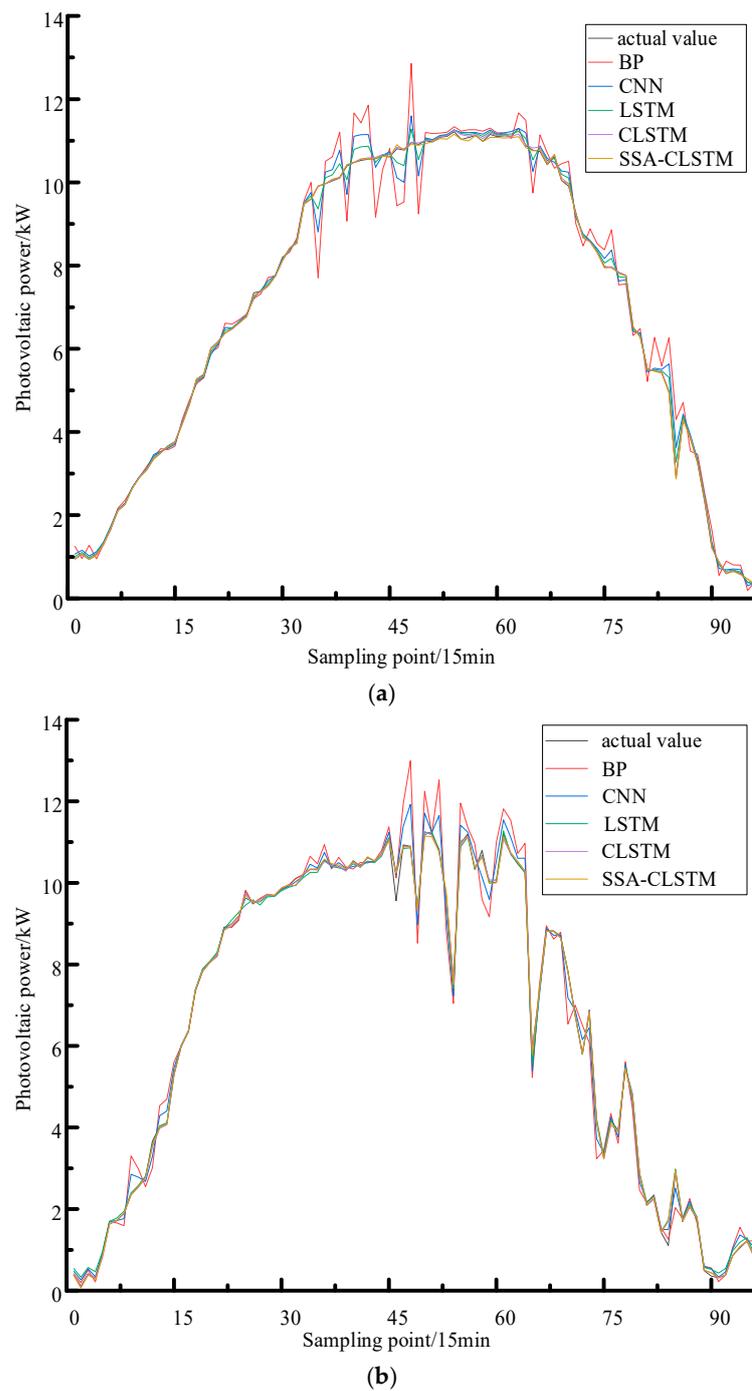
Figure 9. The prediction results of different prediction methods under the two weather types. (**a**) Forecast results of slow weather changes; (**b**) Forecast results of severe weather changes.

Table 1. Errors of different prediction models under two weather types.

Predictive Model	MAE/%		MAPE/%		MSE/%		RMSE/%	
	a	b	a	b	a	b	a	b
BP	12.33	18.56	10.91	17.16	1.34	3.05	16.34	19.24
CNN	7.68	15.37	7.59	14.92	0.98	1.97	12.11	15.03
LSTM	6.08	9.16	4.20	8.52	0.84	1.76	9.18	14.32
CLSTM	3.28	6.43	1.36	2.84	0.39	1.03	6.26	10.28
SSA-CLSTM	2.92	4.29	1.02	2.19	0.34	0.94	1.36	4.71

In terms of convergence speed, it needs to be particularly pointed out that, due to the improvement of the Tent chaotic map, the population initialization is more reasonable and the accuracy of the approximate solution is improved. At the same time, according to the SSA, the iterative process can be updated to the location of the optimal solution faster, and the convergence speed is very fast, as shown in Figure 10. The optimized hybrid network model not only converges faster, but also slightly improves prediction accuracy. Even if the convergence speed is faster, it is not easy to fall into the local optimal situation under the action of the improved chaotic mapping. It is precisely because the improved SSA has the above advantages that it will show excellent performance in the prediction results of several comparison models.
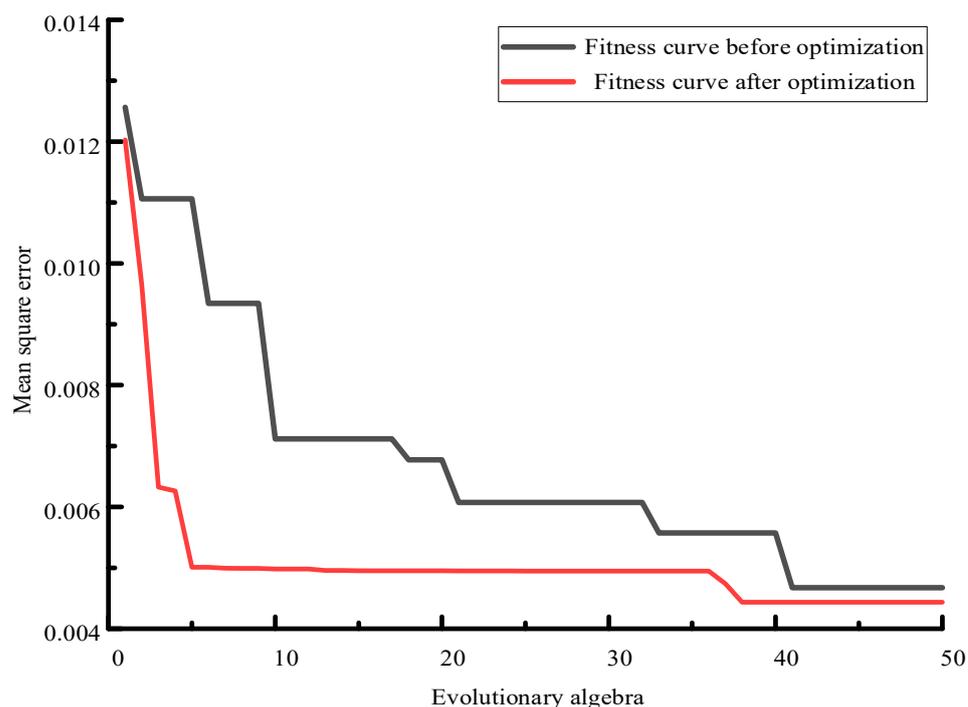

Figure 10. Convergence curve of the hybrid network model.

6.5. Discussion

When designing the method, this paper mainly considered the improvement of prediction accuracy of the CLSTM hybrid neural network model and the performance of weather adaptability. However, there are some parameters, which result in calculation burden in training the model, that need to be optimized. Additionally, although the improved variational modal decomposition can be used to adaptively determine the number of modal components, the residual amount after decomposition is still ignored and original information is partially lost, which will inevitably cause error to be difficult to reduce from the root cause. Therefore, it is the focus of our future work to reduce the time consumed in model training and the degree of loss of original information.

7. Conclusions

Under the development trend of a new-type power system with a high proportion of renewable energy, existing research methods cannot meet the requirements of power system operation and regulation for high accuracy and speed of high proportion photovoltaic output prediction. A combined prediction model of photovoltaic output that optimizes the parameters of the CLSTM hybrid neural network by sparrow search algorithm is proposed. Accurate prediction of photovoltaic output can provide powerful data support for its smoothing and consumption and reduction of its impact on the grid, which is essential

for formulating economic and reliable power dispatch plans. Through simulation and comparative analysis, the following conclusions can be drawn:

(1) The improved variational modal decomposition can adaptively determine the modal quantity parameter, *K*, which can avoid using the cumbersome center frequency domain method to determine *K*.

(2) The proposed fluctuation feature recognition model can divide historical photovoltaic power sequence samples into two types, slow weather changes and severe weather changes. After improved variational modal decomposition technology is used to decompose different types of weather changes, the modal component fluctuation forms have fewer components and the regularity of fluctuations is also improved, compared with the component fluctuation forms obtained by the model without the fluctuation feature recognition.

(3) Since LSTM is better than CNN in dealing with more complex problems, the combination of CNN and LSTM methods, in which the high-frequency component is trained with LSTM, and the low-frequency component is trained with CNN, retains the advantage of CNN with fewer parameters to be determined, while still giving full play to the advantages of LSTM in handling complex problems.

(4) The hybrid neural network model optimized by SSA improves the gradient disappearance or explosion problem in the training process of the traditional optimization algorithms. The advantages of the SSA can also avoid falling into the local optimum, slow iteration speed, initial population random and other issues.

(5) The prediction effect of the model proposed in this paper is significantly better than the BP, CNN, and LSTM single network models, and is also better than the unoptimized CLSTM hybrid model.

When designing the method, this paper mainly considered improvement in the prediction accuracy of the CLSTM hybrid neural network model and the performance of weather adaptability. However, there are some parameters that need to be optimized, which result in calculation burden for training the model. Additionally, although the improved variational modal decomposition can be used to adaptively determine the number of modal components, the residual amount after decomposition is still ignored and original information is partially lost, which will inevitably cause error to be difficult to reduce from the root cause. Therefore, it is the focus of our future work to reduce the time consumed in model training and the degree of loss of original information.

Author Contributions: Writing—original draft preparation, Conceptualization, methodology, validation, S.L.; data curation and project administration, J.Y.; resources, writing—review and funding acquisition, F.W.; validation and supervision R.L.; formal analysis and investigation G.I.R. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: This article is not about human research.

Data Availability Statement: The authors chose not to disclose the data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, C.Y.; Duan, Q.Q.; Zhou, K. A hybrid model for photovoltaic power prediction of both convolution and long short-term memory neural networks optimized by genetic algorithm. *Acta Phys. Sin.* **2020**, *69*, 149–155.
2. Shi, K.J.; Zhang, D.X.; Han, X.Q.; Xie, Z.J. Digital twin model of photovoltaic power generation prediction based on LSTM and migration learning. *Power Syst. Technol.* **2021**, *46*, 1363–1372.
3. Ospina, J.; Newaz, A.; Faruque, M.O. Forecasting of PV plant output using hybrid wavelet-based LSTM-DNN structure model. *IET Renew. Power Gener.* **2019**, *13*, 1087–1095. [CrossRef]

4. Zhang, W. Short-Term solar power forecasting and uncertainty analysis using long and short-term memory. *J. Nanoelectron. Optoelectron.* **2021**, *16*, 1948–1955. [CrossRef]

5. Zhang, N.; Ge, L.J. Combination forecast of photovoltaic short-term output interval based on SOA optimization. *Acta Sol. Energy* **2021**, *42*, 252–259.

6. Hui, L.; Ren, Z.Y.; Yan, X.; Li, W.Y.; Bo, H. A Multi-Data Driven Hybrid Learning Method for Weekly Photovoltaic Power Scenario Forecast. *IEEE Trans. Sustain. Energy* **2022**, *13*, 91–100.

7. Rocha, A.S.F.; Guerra, F.K.D.M.V.; Vale, M.R.B.G. Forecasting the Performance of a Photovoltaic Solar System Installed in other Locations using Artificial Neural Networks. *Electr. Power Compon. Syst.* **2020**, *48*, 201–212. [CrossRef]

8. Guan, L.; Zhao, Q.; Zhou, B.R. Multi-scale clustering analysis based modeling of photovoltaic power character-istics and its application in prediction. *Autom. Electr. Power Syst.* **2018**, *42*, 24–30.

9. Chaibi, Y.; El Rhafiki, T.; Simon-Allue, R. Physical models for the design of photovoltaic/thermal collector systems. *Sol. Energy* **2021**, *226*, 134–146. [CrossRef]

10. Pan, M.Z.; Li, C.; Gao, R. Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization. *J. Clean. Prod.* **2020**, *277*, 123948. [CrossRef]

11. Zhou, Y.; Zhou, N.; Gong, L. Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine. *Energy* **2020**, *204*, 117894. [CrossRef]

12. Wen, X.Q.; Yang, X. Wind Turbine Power Characteristic Analysis and Predictive Modeling Based on EMD Decomposition. *Acta Sol. Energy* **2021**, *42*, 293–298.

13. Su, W.; Jiang, X.; Zeng, L. Ultra-short-term photovoltaic power prediction based on VMD-DESN-MSGP model. *Power Syst. Technol.* **2020**, *44*, 917–926.

14. Nie, Y.; Sun, Y.; Chen, Y. PV power output prediction from sky images using convolutional neural network:The comparison of sky-condition-specific sub-models and an end-to-end model. *J. Renew. Sustain. Energy* **2020**, *12*, 046101. [CrossRef]

15. Zhang, Q.; Ma, Y.; Li, G.L. Applications of frequency domain decomposition and deep learning algorithms in short-term load and photovoltaic power forecasting. *Proc. CSEE* **2019**, *39*, 2221–2230.

16. Lu, J.X.; Zhang, Q.P.; Yang, Z.H. Short-term load forecasting method based on CNN-LSTM hybrid neural network model. *Autom. Electr. Power Syst.* **2019**, *43*, 131–137.

17. Zhou, F.C.; Tang, G.J.; He, Y.L. Unbalanced fault feature extraction for wind power gearbox based on improved VMD. *Vib. Shock* **2020**, *39*, 170–176.

18. Ye, L.; Pei, M.; Lu, P. Combination forecasting method of short-term photovoltaic power based on weather classification. *Autom. Electr. Power Syst.* **2021**, *45*, 44–54.