energies

MDPI

Article

Using Crested Porcupine Optimizer Algorithm and CNN-LSTM-Attention Model Combined with Deep Learning Methods to Enhance Short-Term Power Forecasting in PV Generation

Yiling Fan [1,2,*], Zhuang Ma [1,2], Wanwei Tang [1,2], Jing Liang [1,2] and Pengfei Xu [1,2]

[1] Hebei Key Laboratory of Intelligent Data Information Processing and Control, Tangshan University, Tangshan 063000, China; t0499@tsc.edu.cn (Z.M.); ww_tang@tsc.edu.cn (W.T.); zoey_lj@163.com (J.L.); t1559@tsc.edu.cn (P.X.)

[2] Tangshan Key Laboratory of Intelligent Motion Control System, Tangshan University, Tangshan 063000, China

* Correspondence: fyl@tsc.edu.cn

Abstract: Due to the inherent intermittency, variability, and randomness, photovoltaic (PV) power generation faces significant challenges in energy grid integration. To address these challenges, current research mainly focuses on developing more efficient energy management systems and prediction technologies. Through optimizing scheduling and integration in PV power generation, the stability and reliability of the power grid can be further improved. In this study, a new prediction model is introduced that combines the strengths of convolutional neural networks (CNNs), long short-term memory (LSTM) networks, and attention mechanisms, so we call this algorithm CNN-LSTM-Attention (CLA). In addition, the Crested Porcupine Optimizer (CPO) algorithm is utilized to solve the short-term prediction problem in photovoltaic power generation. This model is abbreviated as CPO-CLA. This is the first time that the CPO algorithm has been introduced into the LSTM algorithm for parameter optimization. To effectively capture univariate and multivariate time series patterns, multiple relevant and target variables prediction patterns (MRTPPs) are employed in the CPO-CLA model. The results show that the CPO-CLA model is superior to traditional methods and recent popular models in terms of prediction accuracy and stability, especially in the 13 h timestep. The integration of attention mechanisms enables the model to adaptively focus on the most relevant historical data for future power prediction. The CPO algorithm further optimizes the LSTM network parameters, which ensures the robust generalization ability of the model. The research results are of great significance for energy generation scheduling and establishing trust in the energy market. Ultimately, it will help integrate renewable energy into the grid more reliably and efficiently.

Keywords: photovoltaic; time series; MRTPP; CPO; attention mechanism

check for updates

Citation: Fan, Y.; Ma, Z.; Tang, W.; Liang, J.; Xu, P. Using Crested Porcupine Optimizer Algorithm and CNN-LSTM-Attention Model Combined with Deep Learning Methods to Enhance Short-Term Power Forecasting in PV Generation. *Energies* **2024**, *17*, 3435. https://doi.org/10.3390/en17143435

Academic Editors: Polo Jesús and López Rodríguez Gabriel

Received: 30 May 2024
Revised: 6 July 2024
Accepted: 9 July 2024
Published: 12 July 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

PV energy emerges as a leading renewable energy source, boasting abundant resources, accessibility, and low operational costs. With the escalating energy demand in developing countries, the growth rate of PV power generation has skyrocketed, posing significant challenges to the electrical system capacity [1]. The International Energy Agency's 2023 report [2] highlights a notable increase in global PV generation in 2022, with the total installed capacity reaching 1185 gigawatts and an addition of 240 gigawatts. PV systems have significantly reduced CO_2 emissions from electricity by approximately 1399 tons, underscoring their role in decreasing both electricity costs and emissions. Nonetheless, the intermittent, volatile, and unpredictable output of grid-connected PV systems frequently disrupts the operation, dispatch, and planning of power systems [3]. Precise PV power forecasting markedly improves solar energy utilization, boosting power stations' return on

investment and minimizing economic losses due to power constraints [4]. Consequently, research into PV power forecasting is crucial and holds substantial value.

PV forecasting methods are categorized into two primary types: statistical and artificial intelligence (AI) models. Statistical models, including the Autoregressive Moving Average (ARMA) [5], Autoregressive Integrated Moving Average (ARIMA), and Seasonal Autoregressive Integrated Moving Average (SARIMA) [6], have been employed. While these models effectively process time series data, they are less adept at managing nonlinear and high-dimensional datasets [7]. With the advancement of artificial intelligence, newer methods have demonstrated superior capabilities in addressing these challenges. These AI models excel at precisely identifying the dynamic properties of PV power generation [8]. They autonomously learn and discern complex data patterns, enhancing their efficacy in forecasting PV output. This evolution marks a significant shift, suggesting a promising future for AI in PV forecasting [9].

Deep learning indeed showcases remarkable abilities in generalization and automatic feature extraction [10]. Agga et al. [11] conducted experiments comparing common deep learning prediction models with machine learning prediction models, confirming the superior effectiveness of deep learning methods. Hence, in predicting PV power, deep learning methods [12] like Artificial Neural Networks (ANNs), CNNs, and Recurrent Neural Networks (RNNs) are typically selected as base models and combined with other methods to create hybrid models for improved accuracy. Zhang et al. [13] proposed a Wavelet Neural Network (WNN) model based on the Genetic Algorithm (GA) and validated it using experimental data collected every half hour under four typical weather conditions at PV power stations. Feng et al. [14] introduced a novel hybrid model called KS-CEEMDAN-SE-LSTM, which captures similar characteristics of PV power generation by decomposing and reconstructing data to reduce non-stationary and noisy features. Case studies have demonstrated its strong performance in short-term PV power prediction. Trong et al. [15] employed Variational Mode Decomposition (VMD) for data preprocessing and proposed a new approach for short-term PV power prediction using the Transformer Neural Network (TransNN) and CNN.

The deep learning models mentioned have demonstrated high accuracy and the ability to capture spatio-temporal features in predicting PV power [16], resulting in favorable outcomes. However, some networks encounter the vanishing gradient issue during long-term sequence prediction, which hinders their capability to manage long-term dependencies within the dataset [17]. Additionally, these models often lack interpretability, are susceptible to overfitting, and demand substantial computational resources [18]. The attention mechanism, as an effective information filtering technique, enables the model to focus adaptively on important features by adjusting the weights of multiple input feature vectors, thereby enhancing the weight of important information and reducing the computational resource requirements [19]. Mirza et al. [20] introduced a deep learning model that integrates a transformer architecture, residual networks, and multi-head attention mechanisms. This model introduces attention mechanisms that selectively focus on pertinent information and understand long-term dependencies within the dataset, thereby improving the accuracy of predictions for wind and PV power output. Yin et al. [21] developed a short-term wind power prediction model that integrates an improved attention mechanism into the Inception Embedding Attention Memory Fully Connected Network. This model outperformed all other comparison algorithms (including EfficientNet, NasNet, and ResNet) by more than 40% on all evaluation metrics. Wang et al. [22] utilized the BiGRU-Attention model to improve the prediction accuracy of low-frequency sequences in wind power. They then aggregated the predicted values from all components to generate the final prediction outcome. These studies collectively showcase the ability of self-attention mechanisms to extract significant features and identify temporal patterns in data, highlighting their significant potential for application in new-energy prediction.

Due to the fluctuation, randomness, and temporality of meteorological data in PV power prediction [23], RNNs and LSTM, which have advantages in handling time series

data in deep learning methods, are often used for the temporal feature analysis of PV data, while convolutional neural networks (CNNs) are also used for the spatial feature analysis of PV data because of their strong feature extraction capabilities [24]. However, these models require hyperparameter tuning, and the choice of parameters determines the quality of the model's prediction results, often using bio-inspired optimization algorithms to identify the optimal hyperparameters for the model [25]. The Crested Porcupine Optimizer (CPO) [26] algorithm, proposed in 2023 by Abdel-Basset et al., is a nature-inspired metaheuristic method that utilizes the visual, auditory, olfactory, and physical attack defense mechanisms of the crested porcupine, corresponding to the algorithm's exploration and exploitation behaviors, to precisely optimize large-scale problems. The algorithm introduces a technique for reducing the cyclic population, activating the defense mechanism only for threatened individuals, thereby improving the convergence speed and population diversity. It exhibits a significantly superior performance on most test functions in the CEC2014, CEC2017, and CEC2020 benchmark tests. Currently, the theory and data of the CPO algorithm are sufficient, but it has not yet been applied to model optimization problems.

In this study, our primary focus revolves around three key aspects. Firstly, we delved into the CNN-LSTM PV power prediction model, leveraging deep learning technology and incorporating an attention mechanism to dynamically extract crucial features. Secondly, we performed a comparative analysis of the predictive performances between models with and without the attention mechanism, alongside the introduction of the CPO algorithm to optimize LSTM parameters. Lastly, we tested the algorithm for embedding MRTPP patterns with different step sizes and compared its performance with those of other algorithms. Furthermore, this paper delineates the following key contributions:

(1) The introduction of a deep learning model, CNN-LSTM, that incorporates an attention mechanism. By leveraging this model, we can fully extract the spatio-temporal changing features of parameters, enabling the CLA model to effectively focus on crucial historical data for future power prediction, thus enhancing the prediction performance.
(2) To enhance the model's predictive ability further, we integrated the CPO algorithm to more efficiently adjust LSTM network parameters, resulting in the formation of the CPO-CNN-LSTM-Attention model. Notably, this is the first instance where the CPO algorithm has been utilized for parameter optimization in the LSTM algorithm, to the best of our knowledge.
(3) Experimental findings suggest that the proposed PV power prediction model surpasses other classical models in accuracy, demonstrating promising application prospects.

2. Model Construction

2.1. CNN-LSTM-Attention

2.1.1. CNN

A CNN is a widely used feedforward neural network [27], mainly composed of convolutional layers, pooling layers, and fully connected layers. The number of layers can be adjusted according to the model's needs. The core idea of a CNN is to use convolutional operations to handle data in Euclidean space, thereby offering substantial advantages in time series prediction [28].

In CNNs, the convolutional layers are mainly used to extract time series features, with advantages such as local perception, weight sharing, and spatial invariance. The pooling layer is used to reduce the dimensionality of the data after convolutional operations to prevent overfitting. Common methods include max pooling and average pooling. The fully connected layer maps the extracted features and passes them to the classifier for classification or regression processing.

This study employed a 1D CNN with convolutional kernels for time series prediction, as depicted in Figure 1. The goal was to capture short-term sequence pattern features and share parameters, consequently reducing the number of parameters necessary for model optimization and computational complexity. This approach enhances the model's training efficiency and scalability.

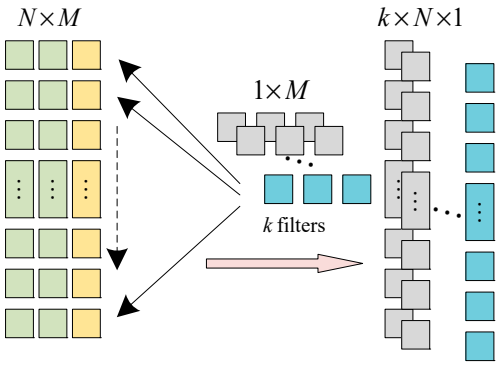


Figure 1. One-dimensional CNN extracting features from time series.

2.1.2. LSTM

LSTM networks are developed based on traditional RNNs, introducing gating mechanisms to help mitigate the issues of vanishing and exploding gradients that RNNs face when managing long-term dependencies [29]. The network structure of an LSTM model includes three gate units: the forget gate, the input gate, and the output gate. The state transition at each timestep depends not only on the previous state but also on the joint influence of these three gate units [30].

The forget gate plays a crucial role in controlling the extent of information transferred from the preceding state to the current state. Concurrently, the input gate modulates the influence of newly received information on the present state, and the output gate governs the dependency of the current output on the state of the memory cell. When the value of the input gate approaches 0 while the value of the forget gate approaches 1, the LSTM network achieves long-term memory functionality. The network filters and retains old state information without adding new input information. Conversely, when the value of the input gate approaches 1 while the value of the forget gate approaches 0, the LSTM network achieves memory updating functionality. The network ignores old information that is not important for the current task and focuses on the current input information.

The collaborative mechanism between gates makes LSTM more flexible in learning and memory processes, allowing it to adapt to different time scales. Therefore, in tasks involving the processing and prediction of time series data, LSTM often outperforms the traditional RNN.

Figure 2 shows the structure of an LSTM unit, with Equations (1)–(6):

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \tag{1}$$

$$f_t = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right) \tag{2}$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \tag{3}$$

$$\widetilde{C}_t = \tanh(W_c x_t + U_c h_{t-1}) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{C}_{t-1} \tag{5}$$

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

where i_t, f_t, and o_t, respectively, correspond to the switch states of the input gate, forget gate, and output gate; $\widetilde{c}_t$ is the candidate cell state; c_t is the current cell state; h_t is the current hidden state; x_t is the value of the input sequence at the current timestep; W and

U represent the weight matrices of the three gates; b represents the bias vector; and σ represents the sigmoid activation function, while tanh denotes the hyperbolic tangent activation function.

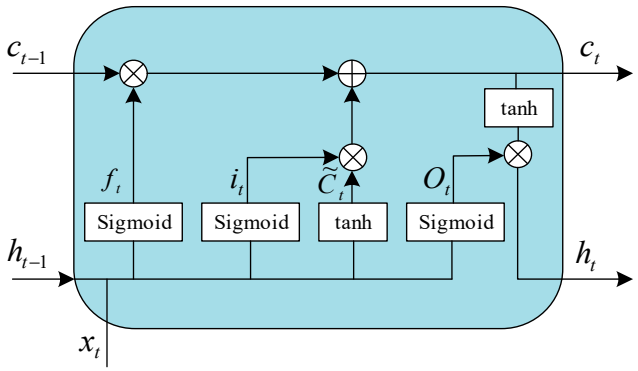


Figure 2. Structure of LSTM unit.

2.1.3. Attention

The CNN-LSTM model has been widely used in PV research. However, its performance is limited by computational power and optimization algorithms [31]. Models can overcome these limitations by introducing attention mechanisms, which mimic the human brain's information processing and significantly improve neural networks' ability to handle spatio-temporal data [32]. This leads to better generalization and optimized model performances.

Attention can be conceptualized as a weighted summation, where the weights correspond to the similarity between the calculation vectors. The original attention includes Q, K, and V, which originate from the input features themselves. V represents the vector of input features, and Q and K are the feature vectors used to calculate the attention weights. When an attention network is not introduced, only one set of V needs to be input into the network for training. However, when an attention network is introduced, this set of V needs to be multiplied by a set of weights $F(Q, K)$, enabling the network to focus on local input features.

Figure 3 shows the structure of the attention. Firstly, attention scores are obtained by calculating the similarity or correlation between a query and keywords (Key), with common methods including the dot product, scaled dot product, and additive attention. Then, the softmax function is used to normalize the attention scores, converting them into a probability distribution that represents the importance of each value. Finally, the values are weighted and summed according to the normalized attention scores using a weighted sum approach (the * in Figure 3 represents multiplication), resulting in a weighted average representation that highlights information more relevant to the query. Equations (7)–(9) are used to obtain these parameters:

$$s_i = F(Q, k_i) \tag{7}$$

$$\alpha_i = softmax(s_i) = \frac{\exp(s_i)}{\sum_{j=1}^{N} \exp(s_j)} \tag{8}$$

$$Attention((K, V), Q) = \sum_{i=1}^{N} \alpha_i v_i \tag{9}$$

The calculated attention values can predict the final PV power output by computing the weighted sum through a fully connected network.

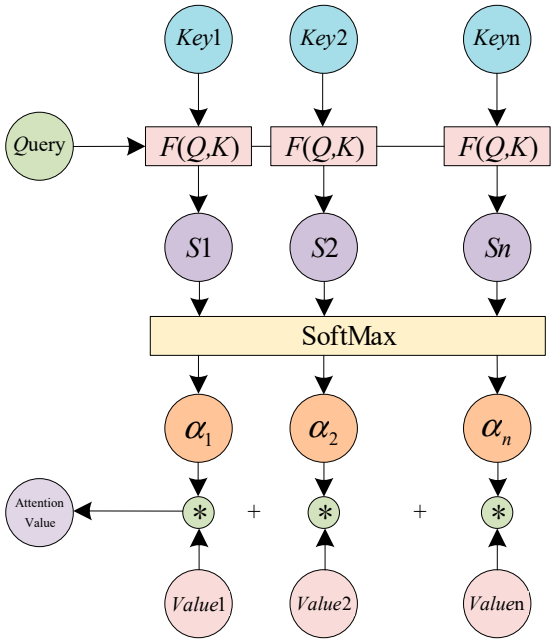


Figure 3. Attention structure.

2.2. CPO

LSTM is susceptible to overfitting when processing sequential data, particularly with limited data [33]. Overfitting diminishes the model's ability to generalize, necessitating meticulous hyperparameter tuning for optimal performance. However, this procedure is time-consuming and might result in excessive tuning, potentially compromising the model's stability and reliability [34]. Thus, when designing LSTM models, it is essential to incorporate optimization algorithms to address overfitting and parameter adjustment, ensuring the model's effectiveness and robustness [35].

The CPO, a novel metaheuristic algorithm proposed in 2023, mimics the defensive behavior of crested porcupines in nature for parameter optimization. When faced with threats, crested porcupines adopt four different defense mechanisms—visual, auditory, olfactory, and physical attack—depending on the severity of the threat. The CPO algorithm maps these defense behaviors into the optimization problem-solving process, where the first two (visual and auditory) are considered exploration strategies, while the latter two (olfactory and physical attack) are considered exploitation strategies.

To improve the algorithm performance, particularly for large-scale optimization problems, the CPO algorithm introduces a new strategy called "cyclic population reduction technology". This technology mimics the behavior of crested porcupines, which activate their defense mechanisms only when directly threatened. This innovation enables the algorithm to expedite convergence while preserving population diversity. This enhancement not only boosts the efficiency and accuracy of the CPO algorithm in tackling optimization problems but also enriches the theoretical frameworks of metaheuristic algorithms. It offers fresh perspectives and methodologies for research in related fields.

Set parameters N', $T_{\max}$, α, T_f, T, and $N_{\min}$, and randomly initialize the population. When $t < T_{\max}$, evaluate the fitness values of the candidate solutions to determine the best solution. Update Equation (10) for defense factor γ_t.

$$\gamma_t = 2 \times rand \times \left(1 - \frac{t}{t_{\max}}\right)^{\frac{t}{t_{\max}}} \tag{10}$$

Update the population size using mathematical model Equation (11) for the cyclic population reduction technology:

$$N = N_{\min} + (N' - N_{\min}) \times \left(1 - \left(\frac{t\%\frac{T_{\max}}{T}}{\frac{T_{\max}}{T}}\right)\right) \tag{11}$$

where N' denotes the population size, α denotes the convergence speed factor, $T_f \in (0,1)$ denotes a predefined constant balancing local exploitation (the third defense mechanism) and global exploitation (the fourth defense mechanism), T denotes a variable determining the number of cycles, t denotes the current function evaluation, $T_{\max}$ denotes the maximum number of function evaluations, % denotes the remainder or modulus operator, and $N_{\min}$ denotes the minimum number of individuals in the newly generated population; thus, the population size cannot be less than $N_{\min}$.

When $i \in (0,1)$, update S and $\vec{\delta}$ and engender two random numbers τ_8 and τ_9; if $\tau_8 < \tau_9$, enter the exploration phase and engender two random numbers τ_6 and τ_7; if $\tau_6 < \tau_7$, engage the first defense mechanism, Equation (12); otherwise, engage the second defense mechanism, Equation (13). If $\tau_8 > \tau_9$, proceed to the development phase and engender a random number τ_{10}; if $\tau_{10} < T_f$, engage the third defense mechanism, Equation (14); otherwise, engage the fourth defense mechanism, Equation (15). Iterate over t to obtain the global optimal fitness value, until $t = T_{\max}$.

$$\vec{x_i^{t+1}} = \vec{x_i^t} + \tau_1 \times \left|2 \times \tau_2 \times \vec{x_{CP}^t} - \vec{y_i^t}\right| \tag{12}$$

$$\vec{x_i^{t+1}} = \left(1 - \vec{U_1}\right) \times \vec{x_i^t} + \vec{U_1}\left(\vec{y} + \tau_3 \times \left(\vec{x_{r1}^t} - \vec{x_{r2}^t}\right)\right) \tag{13}$$

$$\vec{x_i^{t+1}} = \left(1 - \vec{U_1}\right) \times \vec{x_i^t} + \vec{U_1} \times \left(\vec{x_{r_1}^t} + S_i^t \times \left(\vec{x_{r_2}^t} - \vec{x_{r_3}^t}\right) - \tau_3 \times \vec{\delta} \times \gamma_t \times S_i^t\right) \tag{14}$$

$$x_i^{t+1} = \vec{x_{CP}^t} + (\alpha(1 - \tau_4) + \tau_4) \times \left(\delta \times \vec{x_{CP}^t} - \vec{x_i^t}\right) - \tau_5 \times \delta \times \gamma_t \times \vec{F_i^t} \tag{15}$$

where $\vec{x_{CP}^t}$ signifies the best solution obtained and represents CP, $\vec{y_i^t}$ signifies the vector generated between the current CP and a randomly selected CP from the population representing the position of the predator at iteration t, $\vec{U_1}$ signifies a binary vector within the range [0, 1], S_i^t signifies the scent diffusion factor, $\vec{\delta}$ signifies a parameter used to control the search direction, τ_i signifies a random number within [0, 1], $\vec{x_i^t}$ signifies the position of the i-th individual at iteration t and also represents the predator at that position, $\vec{F_i^t}$ signifies the average force affecting the i-th predator's CP, and r_i signifies a random number within [1, N].

2.3. PV Power Forecasting Model

PV power is influenced by external environmental factors, particularly irradiance, temperature intensity, and humidity, which closely correlate with trends in PV power changes [36]. The existence of correlations provides a basis for incorporating relevant variables as input variables into the model. In considering the input–output forecasting models for relevant variables, these encompass the single-input single-output mode for the target variable, the multi-input multi-output mode for both related variables and the target variable, and the multi-input single-output mode for related variables and a single-target variable [37]. To fully utilize multivariate time series data for prediction, MRTPP is proposed, combining the advantages of these three models. In this model, the target variable and its related multivariate time series are used as inputs. In PV power prediction,

the inputs include not only historical data of the PV power itself but also additional factors influencing PV power generation, such as radiation and temperature. The output of the model is the predicted value of the target variable, which is the PV power.

Common strategies for forecasting in multi-step time series encompass direct multi-step prediction and recursive multi-step prediction [38]. The direct multi-step prediction strategy can generate multiple-target variable data simultaneously within the same time-frame, thus limiting the correlation between adjacent values [39]. Therefore, using the recursive multi-step prediction strategy only generates the next PV power output, repeatedly in a rolling manner, but it is necessary to be cautious of the accumulation of future prediction errors, as the data become meaningless once they exceed a certain threshold. Figure 4 shows the multi-relevant target variable prediction model based on the recursive multi-step prediction strategy, where R_i is the multiple input of relevant variables, and O is the input of the target variable, assuming a step length of m, which can predict the target variable of the m + 1th data point.

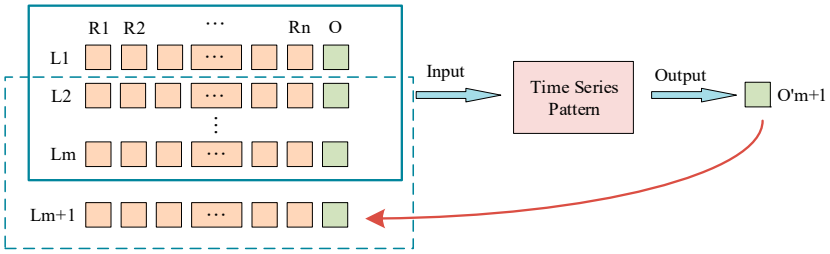


Figure 4. Multi-relevant target variable prediction model based on recursive multi-step prediction strategy.

To obtain experimental results, we initially pinpointed the parameters that needed improvement in the CLA model, encompassing architectural parameters, hyperparameters, and regularization parameters. Subsequently, a fitness function was formulated to assess the CLA model's performance under specific parameter configurations. Any evaluation parameter related to the problem domain, including accuracy and error rate, can serve as the basis for the fitness function. The initial parameter settings of the CPO population are random or predetermined combinations within specified parameter ranges. Through training and analyzing the CLA model with suitable parameter configurations, the fitness of each potential solution in the CPO population was established. The fitness value was calculated based on the model performance evaluated by the fitness function.

In summary, the CLA photovoltaic power prediction model was constructed based on MRTPP, with LSTM parameters optimized using CPO. The entire process is illustrated in Figure 5. The input layer of the model receives time series data, including PV power and other relevant variables. The CNN component first extracts local features through convolution operations, followed by batch normalization and ReLU activation, and then performs max pooling. Convolution is then applied again to extract higher-level features. These feature maps are subsequently fed into the LSTM component, which consists of two LSTM units that capture long-term dependencies in the time series. The CPO algorithm further optimizes the LSTM parameters, enhancing the model's adaptability and prediction accuracy. To improve the model's focus on key features, an attention mechanism is introduced, assigning weighted attention to important parts of the input sequence. The output layer then generates predictions of future photovoltaic power.

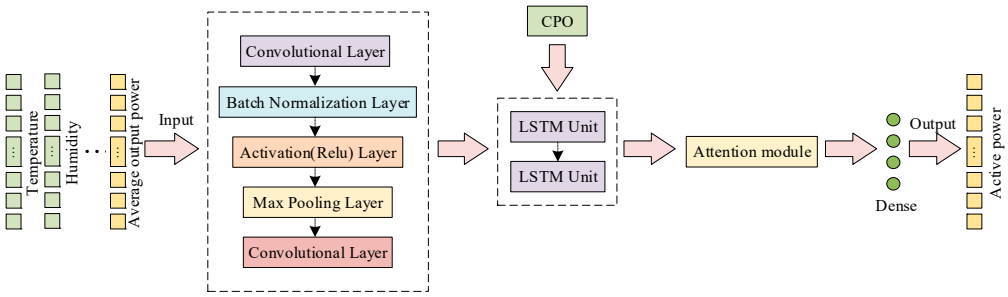


Figure 5. PV power prediction model.

3. Results and Discussion

3.1. Data Collection and Processing

In this study, we evaluated the proposed model (CPO-CLA) using a dataset generated from PV-related features in a certain region, with data monitored hourly in real time on climate and power production between 2020 and 2022 as a case study (see Supplementary Materials). The dataset consisted of 10 different attributes, such as temperature, humidity, pressure, weather id (indicating conditions like sunny, cloudy, rainy), clouds (density of cloud cover), wind speed, wind degree, day of year (reflecting seasonal and daylight changes), clear sky (sky clarity), and average output power. The power output of the PV components is significantly lower in the morning and evening, mostly zero or near zero. Hence, our focus lies solely on the power between 7 a.m. and 7 p.m. Subsequently, the data undergo normalization, outliers are eliminated, and missing values are substituted with the average PV power values from the preceding and succeeding hours. Each dataset timeframe, data volume, percentage, and the model's operating environment are shown in Table 1.

Table 1. Information about data and implementation environment.

Feature	Value
Training data (80%)	1 January 2020–6 May 2022
Testing data (20%)	6 May 2022–30 December 2022
Vector length	10
Sampling rate	1 h
Numerical environment	Python 3.11.0
Libraries	Numpy, Scikit Learn, TensorFlow, Pandas, Scipy
Machine information	12th Gen Intel(R) Core(TM) i7-12700H@2.30 GHz, 64-bit operating system, ×64-based processor

3.2. Objective Function and Evaluation Parameters

The objective function, Mean Squared Error (MSE), is detailed in Equation (16). It calculates the mean of the squared differences between the predicted photovoltaic power output and the actual values of the model's output. In this study, lower values of the objective function are considered more ideal, and minimizing the MSE can enhance the predictive accuracy and model-fitting capabilities. For the model performance assessment, three parameters were employed: the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), both of which serve as prediction bias indicators [40], as detailed in Equations (17) and (18), respectively. The RMSE is highly sensitive to large errors, while the MAE treats each error equally, providing a more direct reflection of the average difference between predicted and actual values. The R-squared value, which quantifies the correlation between actual and predicted values, is presented in Equation (19). In regression analysis,

an R^2 value close to 1 indicates a strong explanatory power of the independent variables over the dependent variable [41].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{16}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{17}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{18}$$

$$R^2 = 1 - \left(\sum_{i}^{n}(y_i - \hat{y}_i)^2\right) / \left(\sum_{i}^{n}\left(y_i - \frac{1}{n}\sum_{i}^{n}y_i\right)^2\right) \tag{19}$$

Here, y_i represents the actual PV power data; $\hat{y}_i$ is the predicted value; and n is the number of test sets.

The training-and-testing procedure of the proposed prediction model is illustrated in Figure 6. First, the dataset is divided into training and testing sets in an 8:2 ratio and normalized. During the training phase, the CPO optimization program is iteratively run to develop new candidate solutions and update each candidate's fitness value using a fitness function, globally searching for better parameter settings. Next, the process checks whether the maximum number of iterations or the highest fitness value has been reached. If these termination conditions are not met, the CPO continues to optimize the CLA hyperparameters. Otherwise, the optimal parameters for the CLA model are updated. In the testing phase, the LSTM global optimal solution found via CPO is used to optimize the CLA model. Finally, the performance of the CLA model on the test dataset is evaluated using the parameters proposed in this section.

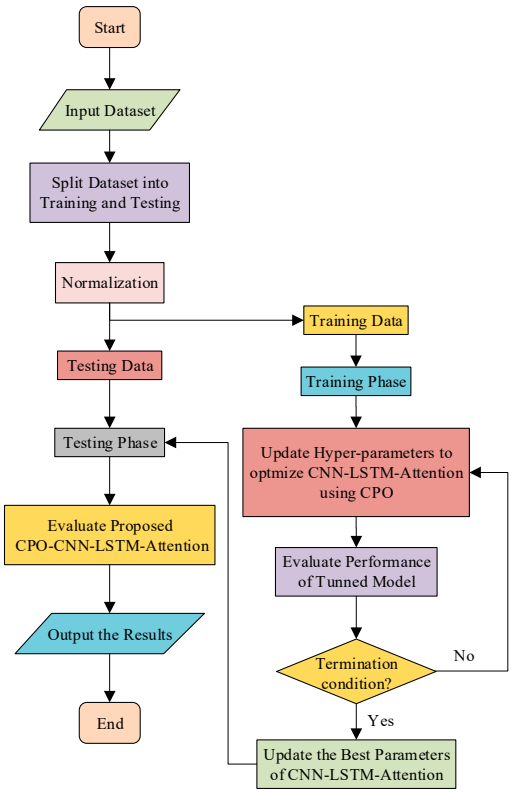


Figure 6. Flowchart of CPO optimizing CLA.

3.3. Prediction Model Result and Evaluation

To assess the predictive capability of the proposed CPO-CLA model, this section conducts a comparative analysis with commonly employed models in PV power forecasting. Before the experiments, the superiority of the CLA model was analyzed by comparing it with LSTM and CNN-LSTM at various timesteps. The details of the model training are presented in Tables 2 and 3. Due to the focus on the data from 7 a.m. to 7 p.m., the natural cycle characteristics in the data can be better captured by selecting 6 and 13 timesteps for model comparison experiments. At 13 timesteps, representing predictions every 13 h helps observe changes over longer time scales, particularly trends and fluctuations across days. At 6 timesteps, representing predictions every 6 h covers different time periods within a day, aiding in capturing the variation patterns of photovoltaic power throughout the day.

Table 2. Training details required for the proposed mode.

Parameters	Details
Epochs	100
Batch size	256
Optimizer	Adam
Learning rate	0.001

Table 3. Parameters of the CLA model.

Parameters		Details
Conv1D	Filter	32
	Kernel size	3
	Activation	ReLu
	Kernel regularizer	L2 (strength 0.1)
MaxPooling1D	pool size	2
Dropout	Dropout Rate	0.3
LSTM	units1	10
	units2	10
Attention	units	20
Dense1	unites	10
	Activation	ReLu
Dense2	unites	1

Table 4 presents the RMSE, MAE, and R^2 scores for the CLA, LSTM, and CNN-LSTM models at 6 and 13 timesteps. Higher R^2 values along with lower RMSE and MAE values indicate superior performance. Figure 7 illustrates the performance regarding statistical data on solar forecasts, encompassing climate and power production data from 2020 to 2022.

Table 4. Performance of the model at 6 and 13 timesteps.

Model	6 Step			13 Step		
	R^2	RMSE	MAE	R^2	RMSE	MAE
CLA	0.868	829.8	670.4	0.875	696.5	601.3
CNN-LSTM	0.793	964.8	718.1	0.754	1053.6	905.3
LSTM	0.512	1494.6	1057.6	0.629	1258.5	953.7

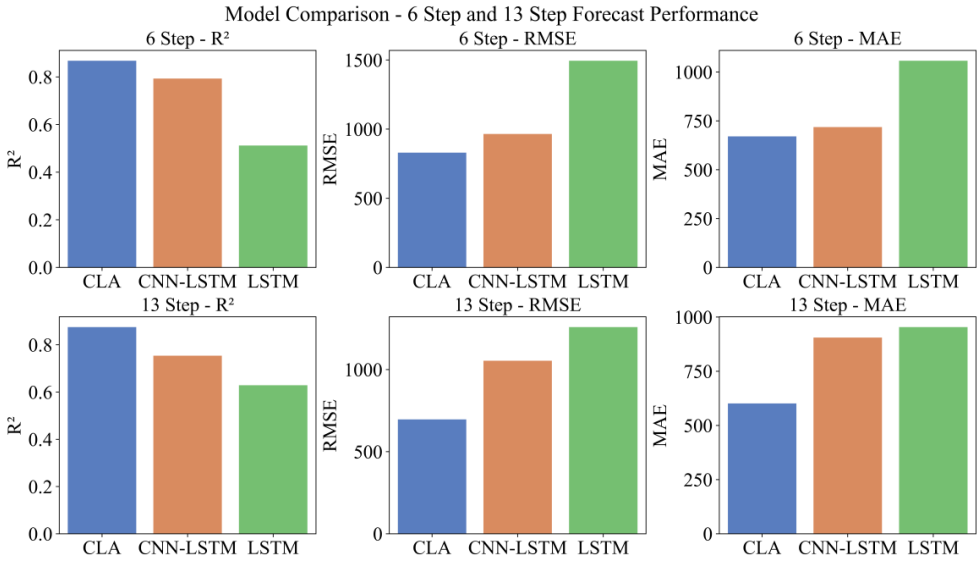


Figure 7. Performance for solar station forecasting.

At a model timestep of 6, the CNN-LSTM significantly outperforms the LSTM model on all three performance metrics, demonstrating that this model can effectively extract spatio-temporal features of PV parameters. In using the CLA approach, the R^2 score is 0.868, the RMSE is 829.8, and the MAE is 670.4, which show a better performance compared to the CNN-LSTM and LSTM models, achieving improvements of 0.075 and 0.356 in the R^2 score, respectively, and also excelling in RMSE and MAE values. When the model timestep is 13, the R^2 value for the CLA model is 0.875, the RMSE is 696.5, and the MAE is 601.3, which exhibit superiority over the CNN-LSTM and LSTM models by 0.121 and 0.246, respectively, in the R^2 value, and even more pronounced improvements in the RMSE and MAE values. In comparing the timesteps of 6 and 13, although the increase in the R^2 value for CLA is only 0.007 and not pronounced, R^2 scores of 0.868 and 0.875 both demonstrate high predictive stability. It is notable that the performance of the CNN-LSTM model decreases at a timestep of 13, which may be due to overfitting during training. Additionally, the longer timestep of 13 better highlights the performance advantage of the attention mechanism in the CLA model, which can adapt well to focus on the most important parts of the historical data for future power prediction, effectively enhancing the predictive performance. This also indicates that the timestep value has a certain impact on the training outcomes of general models, but the CLA model adapts better and is more stable across different time scales.

In the CPO-CLA model, CPO is employed to optimize four parameters of the LSTM: the number of network units, regularizer, and learning rate. Before proceeding, initial parameters must be set to train the CPO. The individual fitness should be updated according to the CP defense strategy, followed by a search for the global optimal solution. Table 5 lists the details of the CPO parameters, Figure 8a illustrates the optimization process of CPO over 50 iterations, and Figure 8b illustrates the training and testing losses of the CPO-CLA model across 100 iterations.

Table 5. CPO parameter settings and optimization range.

Parameters		Details
Pop		3
MaxIter		50
Dim		4
Best parameters	LSTM units1	[16, 128]
	LSTM regularizer	[0.001, 0.01]
	LSTM units2	[16, 64]
	Learning rate	[0.001, 0.01]

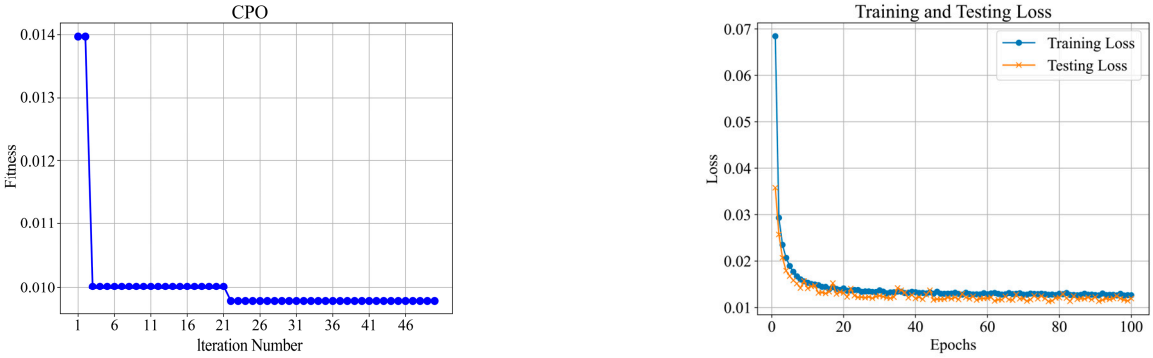


Figure 8. (**a**) CPO optimization process and (**b**) CPO-CLA training and testing process.

Figure 8a demonstrates that the CPO optimization process achieved a fitness value of 0.01 by the third iteration, indicating rapid convergence. It reached an optimum at the 21st iteration, yielding four global optimal solutions for LSTM: 74 units in the first layer, a regularizer of 1.496×10^{-7}, 42 units in the second layer, and a learning rate of 0.0045. Utilizing these parameters, the CLA model was utilized for training and testing with PV forecasting data. The loss curves for the model, as shown in Figure 8b, overlap significantly, reflecting consistent results between training and testing phases. Notably, the model began to converge by the fourth iteration, with a loss value below 0.02, highlighting the CPO's effectiveness in enhancing robustness, simplifying the algorithm, integrating information efficiently, and accelerating convergence.

This analysis confirms that the CPO algorithm optimizes model parameters more effectively. Subsequent comparisons with other optimization algorithms and models are planned to further validate the efficacy of the proposed CPO-CLA model in PV forecasting. Data from the training and testing of various models over 13 timesteps are presented in Table 6, and a comparative analysis of their performance is illustrated in Figure 9.

Table 6. Results for various models at 13 timesteps.

Model	Train			Test		
	R^2	RMSE	MAE	R^2	RMSE	MAE
CPO-CLA	0.974	519.8	331.1	0.965	553.8	360.1
SSA-CLA	0.925	597.1	347.2	0.901	527.5	336.8
SCSO-CLA	0.956	531.9	313.3	0.919	550.1	313.9
CLA	0.874	792.5	598.0	0.857	846.3	641.3
LSTM	0.629	1258.5	852.6	0.616	1206.4	1008.5
CNN-LSTM	0.757	971.3	770.6	0.763	996.3	805.3

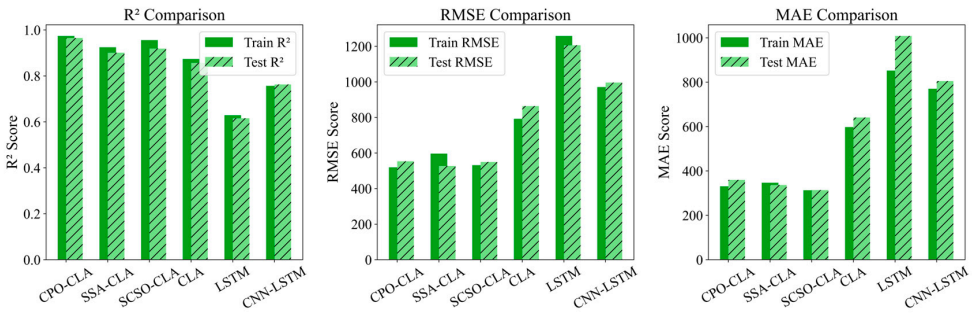
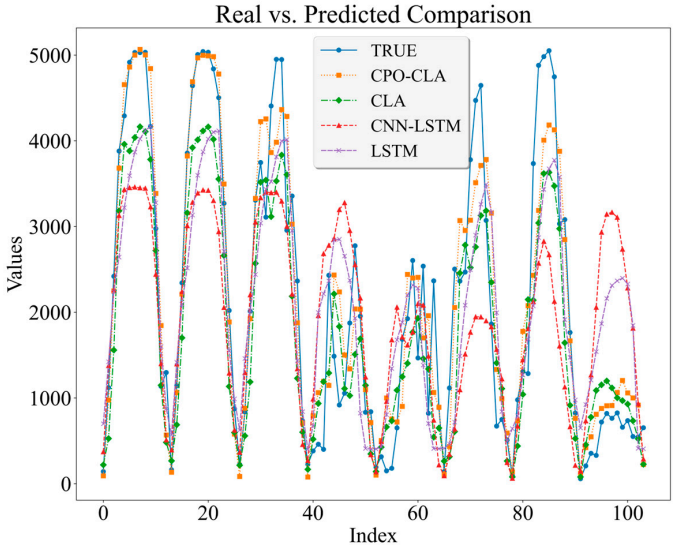


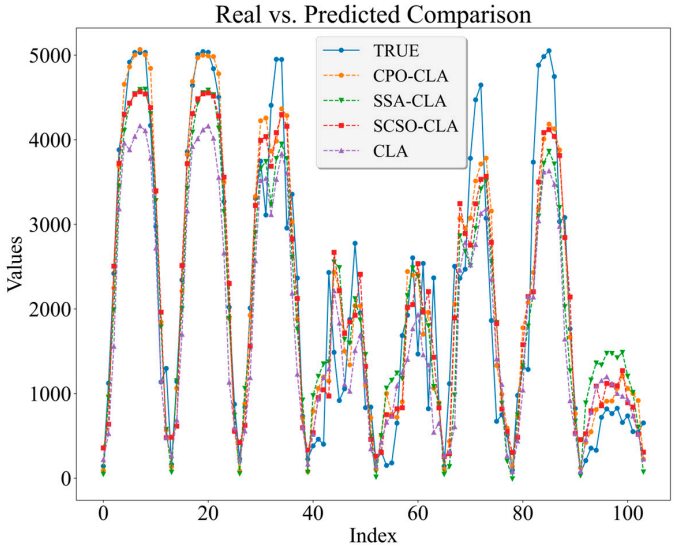
Figure 9. Results for various models at 13 timesteps.

Compared to conventional models, the CPO-CLA model achieves an R^2 score of 0.974, which surpasses the CLA, LSTM, and CNN-LSTM models by 0.1, 0.345, and 0.217,

respectively. When compared to recently developed optimization models, it outperforms the SSA-CLA and SCSO-CLA models by 0.049 and 0.018, respectively, and also registers the lowest scores in the RMSE and MAE. Figure 9 unmistakably illustrates the close alignment between the test findings of the proposed model and the training results, indicating that the CPO algorithm effectively adjusts LSTM parameters to enhance model generalization. Due to the use of a large PV dataset from 2020 to 2022, with a sampling frequency of 1 h, the resulting power prediction curve contains many data points, which are not conducive to display. Therefore, the last 104 data points from the experimental tests are extracted for presentation. Figure 10a compares the prediction effects with conventional models, and Figure 10b compares them with optimized models.



(**a**)



(**b**)

Figure 10. Comparison of predictions with (**a**) conventional models and (**b**) recently developed optimization models.

4. Conclusions

To tackle the challenge of short-term sequence forecasting in PV generation, a novel prediction model, CPO-CLA, has been developed. This model integrates the attention

mechanism from deep learning with the Crested Porcupine Optimizer algorithm into a CNN-LSTM framework, introducing the CPO algorithm into the LSTM algorithm for parameter optimization for the first time. It effectively focuses on the most important parts of historical data using superior time series processing capabilities, enhancing the accuracy and stability of predictions. Additionally, MRTPP is proposed, which leverages both univariate and multivariate time series forecasts, wherein the target variable and its related multivariate time series serve as inputs. The experimental results prove that the CLA model adapts better and more stably across different time scales. The CPO algorithm excels in optimizing model parameters, reducing the loss value to below 0.02 by the fifth iteration, significantly enhancing the model's convergence speed. Compared to traditional methods and recently developed optimization models at a step length of 13, the CPO-CLA model has the lowest RMSE and MAE, with an R^2 score of 0.974, demonstrating superiority and stability, validating its effectiveness and applicability in the day-ahead hourly power prediction of photovoltaic systems. The CPO-CLA model based on MRTPP not only excels in prediction accuracy but also demonstrates consistent performance during model training and testing, reflecting the model's high generalization ability and robustness. However, the model does not predict long-term temporal patterns, limiting its capacity to capture connections between long-term sequence data. Both short-term and long-term temporal patterns could be considered for integration in future forecasting work.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/en17143435/s1.

Author Contributions: Y.F.: data curation (equal); formal analysis (equal); software (equal); validation (equal); and writing—original draft (equal). Z.M.: conceptualization (equal). W.T.: methodology (equal). J.L.: supervision (equal) and visualization (equal). P.X.: writing—review and editing (equal). All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Science Research Project of the Hebei Education Department (grant No. QN2023180).

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

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

References

1. Gaviria, J.F.; Narváez, G.; Guillen, C.; Giraldo, L.F.; Bressan, M. Machine learning in photovoltaic systems: A review. *Renew. Energy* **2022**, *196*, 298–318. [CrossRef]
2. Qi, L.; Song, J.; Wang, Y.; Yan, J. Deploying mobilized photovoltaic system between northern and southern hemisphere: Techno-economic assessment. *Sol. Energy* **2024**, *269*, 112365. [CrossRef]
3. Hussain, A.; Khan, Z.A.; Hussain, T.; Ullah, F.U.M.; Rho, S.; Baik, S.W.; Wei, C. A Hybrid Deep Learning-Based Network for Photovoltaic Power Forecasting. *Complexity* **2022**, *2022*, 7040601. [CrossRef]
4. Liu, W.; Ren, C.; Xu, Y. Missing-Data Tolerant Hybrid Learning Method for Solar Power Forecasting. *IEEE Trans. Sustain. Energy* **2022**, *13*, 1843–1852. [CrossRef]
5. David, M.; Ramahatana, F.; Trombe, P.J.; Lauret, P. Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Sol. Energy* **2016**, *133*, 55–72. [CrossRef]
6. Kushwaha, V.; Pindoriya, N.M. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew. Energy* **2019**, *140*, 124–139. [CrossRef]
7. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-de-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [CrossRef]
8. Wang, M.; Ma, X.; Wang, R.; Kari, T.; Tang, Z. Short-term photovoltaic power prediction model based on hierarchical clustering of K-means++ algorithm and deep learning hybrid model. *J. Renew. Sustain. Energy* **2024**, *16*, 026102. [CrossRef]
9. Abubakar Mas'ud, A. Comparison of three machine learning models for the prediction of hourly PV output power in Saudi Arabia. *Ain Shams Eng. J.* **2022**, *13*, 101648. [CrossRef]
10. Gulay, E.; Sen, M.; Akgun, O.B. Forecasting electricity production from various energy sources in Türkiye: A predictive analysis of time series, deep learning, and hybrid models. *Energy* **2024**, *286*, 129566. [CrossRef]
11. Agga, A.; Abbou, A.; Labbadi, M.; Houm, Y.E.; Ou Ali, I.H. CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. *Electr. Power Syst. Res.* **2022**, *208*, 107908. [CrossRef]

12. Zhang, Y.; Pan, Z.; Wang, H.; Wang, J.; Zhao, Z.; Wang, F. Achieving wind power and photovoltaic power prediction: An intelligent prediction system based on a deep learning approach. *Energy* **2023**, *283*, 129005. [CrossRef]
13. Zhang, C.; Zhang, M. Wavelet-based neural network with genetic algorithm optimization for generation prediction of PV plants. *Energy Rep.* **2022**, *8*, 10976–10990. [CrossRef]
14. Feng, H.; Yu, C. A novel hybrid model for short-term prediction of PV power based on KS-CEEMDAN-SE-LSTM. *Renew. Energy Focus* **2023**, *47*, 100497. [CrossRef]
15. Nguyen Trong, T.; Vu Xuan Son, H.; Do Dinh, H.; Takano, H.; Nguyen Duc, T. Short-term PV power forecast using hybrid deep learning model and Variational Mode Decomposition. *Energy Rep.* **2023**, *9*, 712–717. [CrossRef]
16. An, W.; Zheng, L.; Yu, J.; Wu, H. Ultra-short-term prediction method of PV power output based on the CNN–LSTM hybrid learning model driven by EWT. *J. Renew. Sustain. Energy* **2022**, *14*, 053501. [CrossRef]
17. Feroz Mirza, A.; Mansoor, M.; Usman, M.; Ling, Q. Hybrid Inception-embedded deep neural network ResNet for short and medium-term PV-Wind forecasting. *Energy Convers. Manag.* **2023**, *294*, 117574. [CrossRef]
18. Kumari, P.; Toshniwal, D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl. Energy* **2021**, *295*, 117061. [CrossRef]
19. Yan, Q.; Lu, Z.; Liu, H.; He, X.; Zhang, X.; Guo, J. Short-term prediction of integrated energy load aggregation using a bi-directional simple recurrent unit network with feature-temporal attention mechanism ensemble learning model. *Appl. Energy* **2024**, *355*, 122159. [CrossRef]
20. Mirza, A.F.; Shu, Z.; Usman, M.; Mansoor, M.; Ling, Q. Quantile-transformed multi-attention residual framework (QT-MARF) for medium-term PV and wind power prediction. *Renew. Energy* **2024**, *220*, 119604. [CrossRef]
21. Yin, L.; Zhao, M. Inception-embedded attention memory fully-connected network for short-term wind power prediction. *Appl. Soft Comput.* **2023**, *141*, 110279. [CrossRef]
22. Wang, S.; Shi, J.; Yang, W.; Yin, Q. High and low frequency wind power prediction based on Transformer and BiGRU-Attention. *Energy* **2024**, *288*, 129753. [CrossRef]
23. Abdulai, D.; Gyamfi, S.; Diawuo, F.A.; Acheampong, P. Data analytics for prediction of solar PV power generation and system performance: A real case of Bui Solar Generating Station, Ghana. *Sci. Afr.* **2023**, *21*, e01894. [CrossRef]
24. Agga, A.; Abbou, A.; Labbadi, M.; El Houm, Y. Short-term self consumption PV plant power production forecasts based on hybrid CNN-LSTM, ConvLSTM models. *Renew. Energy* **2021**, *177*, 101–112. [CrossRef]
25. Iweh, C.D.; Akupan, E.R. Control and optimization of a hybrid solar PV—Hydro power system for off-grid applications using particle swarm optimization (PSO) and differential evolution (DE). *Energy Rep.* **2023**, *10*, 4253–4270. [CrossRef]
26. Abdel-Basset, M.; Mohamed, R.; Abouhawwash, M. Crested Porcupine Optimizer: A new nature-inspired metaheuristic. *Knowl. Based Syst.* **2024**, *284*, 111257. [CrossRef]
27. Al-Ja'afreh, M.A.A.; Mokryani, G.; Amjad, B. An enhanced CNN-LSTM based multi-stage framework for PV and load short-term forecasting: DSO scenarios. *Energy Rep.* **2023**, *10*, 1387–1408. [CrossRef]
28. Ghimire, S.; Nguyen-Huy, T.; Deo, R.C.; Casillas-Pérez, D.; Salcedo-Sanz, S. Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. *Sustain. Mater. Technol.* **2022**, *32*, e00429. [CrossRef]
29. Wang, L.; Mao, M.; Xie, J.; Liao, Z.; Zhang, H.; Li, H. Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model. *Energy* **2023**, *262*, 125592. [CrossRef]
30. Ehteram, M.; Afshari Nia, M.; Panahi, F.; Farrokhi, A. Read-First LSTM model: A new variant of long short term memory neural network for predicting solar radiation data. *Energy Convers. Manag.* **2024**, *305*, 118267. [CrossRef]
31. Gao, B.; Huang, X.; Shi, J.; Tai, Y.; Zhang, J. Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks. *Renew. Energy* **2020**, *162*, 1665–1683. [CrossRef]
32. Tovar Rosas, M.A.; Pérez, M.R.; Martínez Pérez, E.R. Itineraries for charging and discharging a BESS using energy predictions based on a CNN-LSTM neural network model in BCS, Mexico. *Renew. Energy* **2022**, *188*, 1141–1165. [CrossRef]
33. Hu, Z.; Gao, Y.; Ji, S.; Mae, M.; Imaizumi, T. Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data. *Appl. Energy* **2024**, *359*, 122709. [CrossRef]
34. Zuo, H.-M.; Qiu, J.; Jia, Y.-H.; Wang, Q.; Li, F.-F. Ten-minute prediction of solar irradiance based on cloud detection and a long short-term memory (LSTM) model. *Energy Rep.* **2022**, *8*, 5146–5157. [CrossRef]
35. Elizabeth Michael, N.; Hasan, S.; Al-Durra, A.; Mishra, M. Short-term solar irradiance forecasting based on a novel Bayesian optimized deep Long Short-Term Memory neural network. *Appl. Energy* **2022**, *324*, 119727. [CrossRef]
36. Qu, J.; Qian, Z.; Pei, Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* **2021**, *232*, 120996. [CrossRef]
37. Liu, J.; Huang, X.; Li, Q.; Chen, Z.; Liu, G.; Tai, Y. Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. *Energy Convers. Manag.* **2023**, *280*, 116804. [CrossRef]
38. Gao, C.; Zhang, N.; Li, Y.; Lin, Y.; Wan, H. Adversarial self-attentive time-variant neural networks for multi-step time series forecasting. *Expert Syst. Appl.* **2023**, *231*, 120722. [CrossRef]
39. El Alani, O.; Abraim, M.; Ghennioui, H.; Ghennioui, A.; Ikenbi, I.; Dahr, F.-E. Short term solar irradiance forecasting using sky images based on a hybrid CNN–MLP model. *Energy Rep.* **2021**, *7*, 888–900. [CrossRef]

40. Ni, Q.; Zhuang, S.; Sheng, H.; Kang, G.; Xiao, J. An ensemble prediction intervals approach for short-term PV power forecasting. *Sol. Energy* **2017**, *155*, 1072–1083. [CrossRef]
41. Jalali, S.M.J.; Osorio, G.J.; Ahmadian, S.; Lotfi, M.; Campos, V.M.A.; Shafie-khah, M.; Khosravi, A.; Catalao, J.P.S. New Hybrid Deep Neural Architectural Search-Based Ensemble Reinforcement Learning Strategy for Wind Power Forecasting. *IEEE Trans. Ind. Appl.* **2022**, *58*, 15–27. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.