WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 4
ICCIA 2024

Research on Photovoltaic Power Generation Prediction Based on Ensemble Learning DBO-LSTM-EBRB

Boyi Fu*, Jiayan Lin[a]

School of Economics and Management, Fuzhou University, Fujian, China

*fuboyi2003@163.com, [a]linjiayan2003@126.com

Abstract. Photovoltaic power generation represents a pivotal form of renewable energy. Accurate power prediction is therefore of paramount importance to the advancement of the energy revolution. This paper examines the Yumara Solar System PV plant in the central region of Australia as a case study. The research begins with the use of the k nearest neighbor algorithm (KNN) and extreme gradient boosting algorithm (XGBoost) to pre-process the PV power generation data from the PV plant. This is followed by the creation of line graphs, which illustrate the changes in variables in the PV power generation data from the PV plant in 2022. Subsequently, this paper employs the dung beetle optimization (DBO) algorithm to optimize the long short-term memory (LSTM) network and construct the DBO-LSTM model. The DBO-LSTM model is further integrated with the extended belief rule base (EBRB) model under the integrated learning framework. The weight assignments are reasonable, in accordance with the model features, for the purpose of constructing the DBO-LSTM-EBRB model for predicting the capacity of photovoltaic power generation. This paper presents experimental results that demonstrate superior predictive performance of the DBO-LSTM-EBRB model in comparison to the traditional LSTM and EBRB models. The R2 value achieved for the DBO-LSTM-EBRB model ranges from 0.950 to 0.978.

Keywords: photovoltaic power generation; power forecasting; LSTM; EBRB; DBO.

1. Introduction

Solar energy is a significant contributor to the field of renewable energy[1]. Accurate power prediction is of great significance to promote the energy revolution[2], as it contributes to improving energy efficiency, optimizing energy usage, and achieving sustainable development in energy[[3],[4]].However, photovoltaic power generation is unstable[5]. Changes in wind speed, temperature and other meteorological factors lead to large fluctuations[6], showing intermittent and random characteristics, which not only brings challenges to power grid dispatching and operation, but also affects the economic benefits and market competitiveness of photovoltaic power generation[7].

Currently, a considerable number of scholars have conducted extensive research on the prediction of photovoltaic power generation. Among these, the prediction method based on artificial intelligence technology is of particular importance. The commonly used methods include multiverse optimization (HIMVO) algorithm, artificial neural network (ANN), gated periodic unit (GRU), neighbor propagation (AP) algorithm, backpropagation (BP) network, etc.

Ma Jun et al.[8] developed a hybrid and improved HIMVO algorithm to optimize the hyperparameters of support vector machines. It is superior to SVM and MVO-SVM under various weather conditions, and has the highest accuracy under sunny conditions, with an average absolute percentage error (MAPE) of only 5.30%. Zhu et al.[9] took meteorological data as input and used ANN model to predict photovoltaic power generation. In the best case, the mean absolute error (MAE) reaches 3.64%, which has a high prediction accuracy. Zhang Jin et al.[10] suggested a GRU-attention model, adding Attention mechanism into the hidden layer of the GRU model to predict accuracy, and its optimal prediction day determination coefficient (R2) was as high as 0.98. Wang Xiaoxia et al.[11] used AP algorithm to cluster photovoltaic power stations, realized concurrent forecast of multiple power stations through short-term time series network (LSTNet), established AP-LSTNET model, and

realized relatively low prediction error. Niu et al.[12] used particle swarm optimization algorithm (DIFPSO) based on dynamic inertia factors to optimize the BP network. Combined with the complementary integrated empirical Mode decomposition (CEEMD) method, the RF-CEEMD-DIFPSO-BPNN model was constructed to decompose the original data sequence of photovoltaic power stations into multiple IMF and residual components for prediction. Following the superimposition of the prediction results for each component in the different typical weather and performance tests of a kilowatt photovoltaic power station for 3 days, the R2 is as high as 0.99, which has a high degree of predictive accuracy.

This paper is based on the real-time monitoring indicators of 11 meteorological characteristics provided by the Yumara Solar System photovoltaic power station in Central Australia in 2022. A photovoltaic power prediction model based on ensemble learning -- Dung Beetle optimization algorithm -- Long short-term memory network -- Extended belief rule base (DBO-LSTM-EBRB) was established to fit the power variation rule of photovoltaic power generation, and provide technical support for improving the performance of photovoltaic power generation.

2. Research method

2.1. Long short-term memory network

The long short-term memory network incorporates a gate structure based on a recurrent neural network (RNN) to store cell states, thereby overcoming the limitations of the gradient explosion and gradient vanishing that are inherent to traditional RNNs. The hidden layer is equipped with a forgetting gate, an input gate, and an output gate, which enables it to perform both long- and short-term memory functions. This capability has a beneficial effect on the processing of time series data.

2.1.1. fogetting gate

The forget gate is used to control whether to discard information in the cell state. First, the output of the hidden layer at t-1 time and the input at the present time are received as input parameters, and then any value within the range of [0,1] is obtained through the activation function, and it is passed to the cell. The corresponding formula is as follows:

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

Where σ is the activation function, w_f stands the weight coefficient matrix between the output h_{t-1} of the hidden layer at time t-1 and the input x_t at the current time to the forgetting gate, and b_f represents the corresponding bias term.

2.1.2. input gate

Input gates are used to control whether the current input information is written to the cell. The input gate first creates the input gate parameter i_t through the σ layer, which is used to determine the updated information;Secondly, new candidate cell states $\tilde{C}_t$ are generated through the *tanh* layer to store the updated candidate values； Then, the obtained forgetting gate parameter f_t is multiplied with the original cell state C_{t-1} to discard part of the information, and the result obtained by multiplying the input gate parameter i_t with the candidate cell state $\tilde{C}_t$ is added to incorporate the updated information, so as to receive the most recent cell state C_t.The corresponding formula is provided below：

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

$$\tilde{C}_t = \tanh\left(W_c \cdot [h_{t-1}, x_t] + b_c\right) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{4}$$

Where w_i and W_c represent the corresponding h_{t-1} and x_t to the input gate activation function σ and *tanh* weight coefficient matrix respectively, b_i and b_c represent the corresponding bias term respectively.

2.1.3. out-gate

Output gates are used to control whether to output information in the cell state. The output gate first normalizes the values in the updated cell state C_t through the *tanh* layer so that they are in the interval [-1,1]. Secondly, the output gate parameter o_t is generated through the σ layer to confirm the message to be output. Then multiply the two parts together to get the final output h_t. The corresponding formula is provided below：

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

$$h_t = o_t \times \tanh\left(C_t\right) \tag{6}$$

Where W_o stands the weight coefficient matrix of h_{t-1} and x_t to the output gate activation function σ, and b_o represents the corresponding offset term.

2.2. Extended belief rule base model

Extended belief rule base (EBRB)[13] reasoning model is based on IF-THEN rules. This model introduces extended belief rule based on the reasoning model of belief rule base (BRB), combines uncertain information with belief, synthesizes evidence on the premise of retaining uncertain information, and the evidence fusion process is clear and transparent. The actual weather conditions of photovoltaic power generation are more complex, and the nonlinear mapping relationship between the influence factors and photovoltaic power generation can be intuitively demonstrated by using EBRB model.

The EBRB model is embedded with the IF-THEN rule of the belief frame, which reflects the "scenario-response" mapping relationship of the system in a specific scenario. k (k = 1,... , L) The extended belief rule is provided below:

$$\begin{gathered} R_k\text{: IF } U_1 \text{ is } \left\{ \left(A_{1,j}, \alpha_{1,j}^k\right); j = 1, \ldots, J_1 \right\} \wedge \cdots \\ \wedge\ U_M \text{ is } \left\{ \left(A_{M,j}, \alpha_{M,j}^k\right); j = 1, \ldots, J_M \right\}, \\ \text{then } D \text{ is } \left\{ \left(D_n, \beta_{n,k}\right); n = 1, \ldots, N \right\}, \\ \text{with rule weight } \theta_k \\ \text{and attribute weights } \{\delta_1, \ldots, \delta_M\} \end{gathered} \tag{7}$$

Among them, $\alpha_{i,j}^k$ represents the belief on the reference level $A_{i,j}$ of the i premise attribute in article k Extended belief rule； $\beta_{n,k}$ stands the belief level assigned to the result attribute D_n in the Extended belief rule k,if $\sum_{n=1}^{N} \beta_{n,k} < 1$,The information contained in Article k Extended belief rule is complete; otherwise, the information contained in Article k Extended belief rule is incomplete； θ_k represents the rule weight of article k belief rule； δ_i represents the attribute weight of the I-th premise attribute.

The detailed modeling steps of EBRB model are provided below:

2.2.1. Initialize the values of key parameters in the EBRB

The expert knowledge is used to give all candidate values in all prerequisite attributes, utility values of all evaluation levels in result attributes and attribute weights of all prerequisite attributes.

2.2.2. Calculate distributed belief

Suppose there are T sets of data, expressed as $<\boldsymbol{x_i}, y_i>(i=1,\ldots, T)$, where x_i stands the input value vector of the i premise attribute; y_i represents the I-th output value; Then, according to the rule based information transformation technique, the distributed belief of the I-th premise attribute and the result attribute is calculated. The corresponding formula is provided below:

$$S\left(x_{k,i}\right) = \left\{ \left(A_{i,j}, \alpha_{i,j}^k\right); j = 1, \cdots, J_i \right\} \tag{8}$$

$$S\left(y_k\right) = \{\left(D_n, \beta_{n,k}\right); n = 1, \cdots, N\} \tag{9}$$

Among then,

$$\alpha_{i,j}^{k} = \frac{u(A_{i,j+1}) - x_{k,i}}{u(A_{i,j+1}) - u(A_{i,j})} \text{ and } \alpha_{i,j+1}^{k} = 1 - \alpha_{i,j}^{k}, \text{ if } u(A_{i,j} \leqslant x_{k,i} \leqslant u(A_{i,j+1})), \quad \alpha_{i,t}^{k} = 0 \text{ for } t = 1, \cdots, J_i \text{ and } t \neq j, j+1 \tag{10}$$

2.2.3. Calculate individual matching degree

The individual matching degree measures the likeness between the input data and the extended belief rule. The higher the likeness, the more consistent the input data is with a rule, and the more likely the rule is to be activated. The degree of matching between the individual premise attribute in Article K and the extended belief rule is provided below:

$$S^{k}(x_i, U_i) = 1 - \sqrt{\frac{\sum_{j=1}^{J_i} (\alpha_{i,j} - \alpha_{i,j}^{k})^2}{2}} \tag{11}$$

2.2.4. Calculate activation weight

The activation weight of the Extended Belief Rule k is obtained by integrating the rule weight, attribute weight, and individual match degree. The corresponding formula is provided below:

$$w_k = \frac{\theta_k \prod_{i=1}^{M} (S^{k}(x_i, U_i))^{\bar{\delta}_i}}{\sum_{l=1}^{L} (\theta_l \prod_{i=1}^{M} (S^{l}(x_i, U_i))^{\bar{\delta}_i})}, \bar{\delta}_i = \frac{\delta_i}{\max_{j=1,\cdots,M} \{\delta_j\}} \tag{12}$$

When $\omega_k > 0$, the k Extended belief rule is activated.

2.2.5. Apply evidential reasoning (ER) algorithm to synthesize belief

All activation rules are synthesized at one time through the analytic formula in the ER algorithm, and the corresponding formula of the synthesized new belief is as follows:

$$\beta_n = \frac{\prod_{k=1}^{L} (\omega_k \beta_{i,k} + 1 - \omega_k \sum_{i=1}^{N} \beta_{i,k}) - \prod_{k=1}^{L} (1 - \omega_k \sum_{i=1}^{N} \beta_{i,k})}{\sum_{i=1}^{N} \prod_{k=1}^{L} (\omega_k \beta_{i,k} + 1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}) - (N-1) \prod_{k=1}^{L} (1 - \omega_k \sum_{j=1}^{N} \beta_{j,k}) - \prod_{k=1}^{L} (1 - \omega_k)} \tag{13}$$

Suppose $u(Dn)$ is the utility value of Dn of the n (n=1,⋯, N) result level, then the predicted value based on EBRB is:

$$f(x) = \{(D_n, \beta_n); n = 1, 2, ..., N\} \tag{14}$$

2.3. Construction of DBO-LSTM-EBRB prediction model

The concept of ensemble learning involves the combination of disparate single algorithms for the purpose of modeling. This approach has been demonstrated to enhance prediction accuracy by integrating the outputs of disparate models. Based on the above related principles, this study integrated learning between the established DBO-LSTM prediction model and EBRB model to build the DBO-LSTM-EBRB prediction model. The modeling process of the DBO-LSTM-EBRB prediction model is displayed in Figure 1.

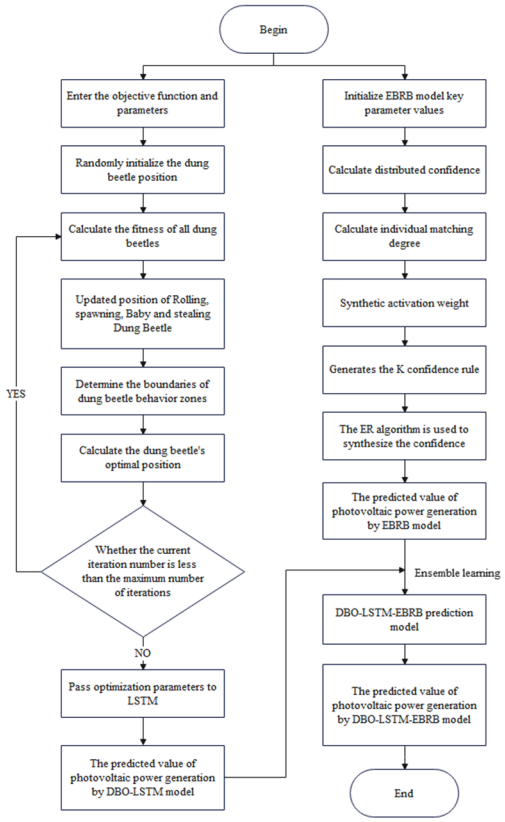

Figure 11. Modeling process of DBO-LSTM-EBRB forecast model

2.4. Model accuracy evaluation method

This paper employs the evaluation metrics of R2, MAE, and root mean square error (RMSE) to assess the efficacy of the proposed model. The R^2 value, which represents the proportion of variance explained by a regression model, ranges from 0 to 1. A value closer to 1 indicates a stronger ability of the independent variable to explain the variance in the dependent variable. MAE is the average absolute discrepancy between the predicted quantity and the true quantity, with the value range being [0, +∞). The closer to 0, the higher the accuracy of the prediction. RMSE is a metric that quantifies the discrepancy between the predicted and actual values. The range of the RMSE value is [0, +∞). A smaller value indicates a smaller prediction error for the model and a stronger prediction ability for the model. The formulas of the above three indexes are shown in equations (15)- (17) :

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y})^2} \tag{15}$$

$$MAE = \frac{1}{m} |y_i - \hat{y}_i| \tag{16}$$

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

3. Experimental verification

3.1. Data source and data preprocessing

This study selected the measured PV power time series data of Yumara Solar System PV power station in 2022. The data set contains the historical power generation data of the photovoltaic power station and the corresponding meteorological data, namely wind direction, weather temperature, wind speed, global horizontal radiation, rainfall, max wind speed, air pressure, hail accumulation amount, pyranometer, temperature probe 1, temperature probe 2, a total of 11 measured data.

Considering the local climate conditions, the data of the amount of hail accumulation is eliminated in this paper. Meanwhile, the temperature probe 1 and temperature probe 2 are integrated into the component temperature by calculating the mean value. Since the effective output period of photovoltaic power is mainly during the day, this paper adopts the data during the period of 7:00-18:00 to carry out the research, the data sampling interval is 5 minutes, the data sampling points are 133 per day, and the total data amount is 48,545.

In this paper, the K-nearest neighbor (KNN) algorithm is employed to supplement the missing values of both independent and dependent variables. By finding K samples that are most similar to the missing samples and predicting the missing values according to their eigenvalues, the box diagram is used to detect outliers, and it is found that there are data beyond the upper bound in the data set, and it is replaced with the upper bound value to eliminate the interference of extreme and anomalous data on the predictive model. Improve the accuracy of prediction results. Finally, the data are normalized to achieve comparability among different impact factors.

3.2. Descriptive statistical analysis

A descriptive statistical analysis of variables and photovoltaic power generation, as presented in Table 1, reveals a clear range of changes in the environmental parameters monitored by photovoltaic power stations. The temperature fluctuates around 24.23℃, the global horizontal radiation value is high, reaching 526.93 W • m^{-2} on average, and the pressure is stable at a high level (956.15hPa on average), while the pyranometer varies widely, the minimum value is negative, and the maximum value can reach 1416.36 W • m^{-2}. The component temperature fluctuates around 50.40℃ and the power averages about 276.86kW.

Table 11. Descriptive statistical analysis

Index	Temperature	Radiation	Air pressure	Pyranometer	Temperature Probe	Photovoltaic power
Sample size	48545	48545	48545	48545	48545	48545
Mean	24.23	526.93	956.15	559.56	50.40	276.86
Minimum	-0.40	-3.86	940.91	-3.60	30.32	0
25%	18.63	237.25	952.33	258.77	41.67	132.68
50%	24.07	525.10	956.07	568.27	49.66	270.71
75%	30.45	791.46	959.90	861.92	58.37	396.34
Maximum	42.91	1407.07	971.58	1416.36	80.12	791.66
MSE	8.16	328.19	5.30	339.30	10.91	178.88

3.3. Correlation analysis

The Pearson correlation coefficient is a statistical measure used to quantify the linear correlation between two variables, expressed as ρ. The value range for this statistic is [-1,1]. The value of ρ is indicative of the strength of the correlation between the two variables in question. The values of -1, 0 and 1 correspond, respectively, to a complete negative correlation, no correlation and a complete positive correlation. The closer the absolute value of ρ is to 1, the stronger the linear correlation between the two variables. In the data set of this paper, there is a linear relationship between different impact factors and photovoltaic power generation, so The Pearson correlation coefficient is employed to quantify the degree of correlation across different impact variables and photovoltaic power generation.

The corresponding formula is as follows:

$$\rho = \frac{E(XY) - E(X)\,E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \tag{18}$$

Where X and Y represent two different variables.

A correlation analysis was conducted to determine the correlation coefficient and the correlation between photovoltaic power and other variables. The findings are depicted in Table 2. As illustrated in the accompanying figure, the positive correlation between rainfall and photovoltaic power generation is relatively weak, with a correlation coefficient of 0.02.The negative correlation between wind direction and photovoltaic power is weak, and the correlation coefficient is -0.085. Therefore, the two variables of rainfall and wind direction are deleted to improve the correlation between the variables and photovoltaic power generation.

Table 22. Correlation coefficient and correlation of PV power with other variables

Feature	Correlation coefficient
Wind speed	0.17
Weather Temperature	0.44
Global Horizontal Radiation	0.57
Wind Direction	-0.085
Rainfall	0.02
Max Wind Speed	0.24
Air Pressure	-0.21
Pyranometer	0.52
Temperature Probe	0.49

3.4. Extreme gradient lifting algorithm feature extraction

The factors that influence the generation of photovoltaic power are numerous and can be reflected in wind speed, wind direction, temperature, air pressure, global horizontal radiation, and other characteristics. To enhance the speed and efficacy of the model training process, the extreme gradient lifting algorithm is employed to identify the most pertinent features within the selected data set. These features are then selected based on their impact on the prediction of PV power, with the historical power generation data and corresponding meteorological data serving as the primary sources of information.

In this study, the extreme gradient lifting algorithm was used to iteratively build a new tree model, and the importance of the obtained features was ranked, as shown in Figure 2. The first 5 influencing factors were selected as independent variables, namely, weather temperature, global horizontal radiation, air pressure, temperature probe and pyranometer, and the corresponding photovoltaic power generation power was taken as the dependent variable to build the model.

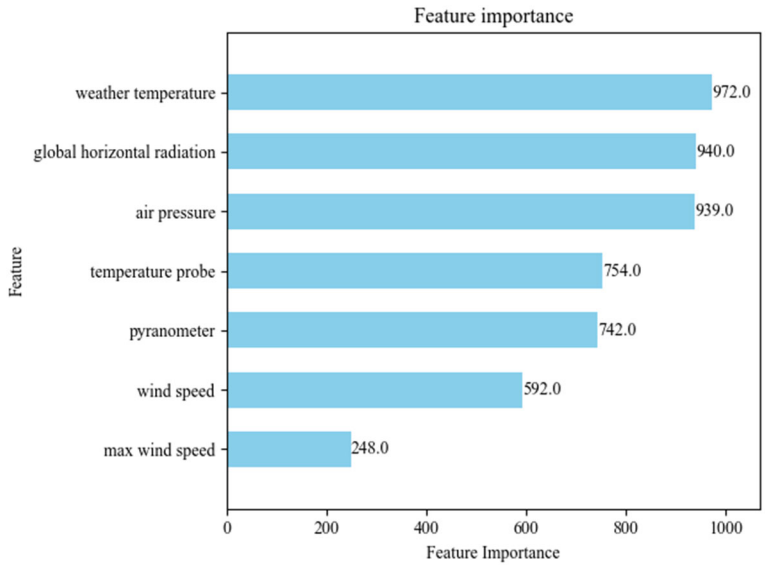

Figure 2. Feature importance result graph

4. Result analysis

4.1. Seasonal analysis

This paper analyzes the daily changes of the mean, maximum and minimum values of each variable in the PV power generation data of Yumara Solar System photovoltaic power station in 2022, and shows the cyclical change trend of these data by drawing a line chart, as shown in Figure 3. The data indicated a clear trend in the temperature of the atmosphere, with a decrease followed by an increase. Similarly, the data revealed a trend in the horizontal radiation, with a decrease followed by an increase. In contrast, the data showed a trend in air pressure, with an increase followed by a decrease. On the other hand, the pyranometer showed a small trend of first decreasing, then rising and then stable fluctuating. The overall change trend of module temperature is more consistent with the change of power, indicating that temperature has a greater impact on power, and directly affects the performance and efficiency of photovoltaic modules.

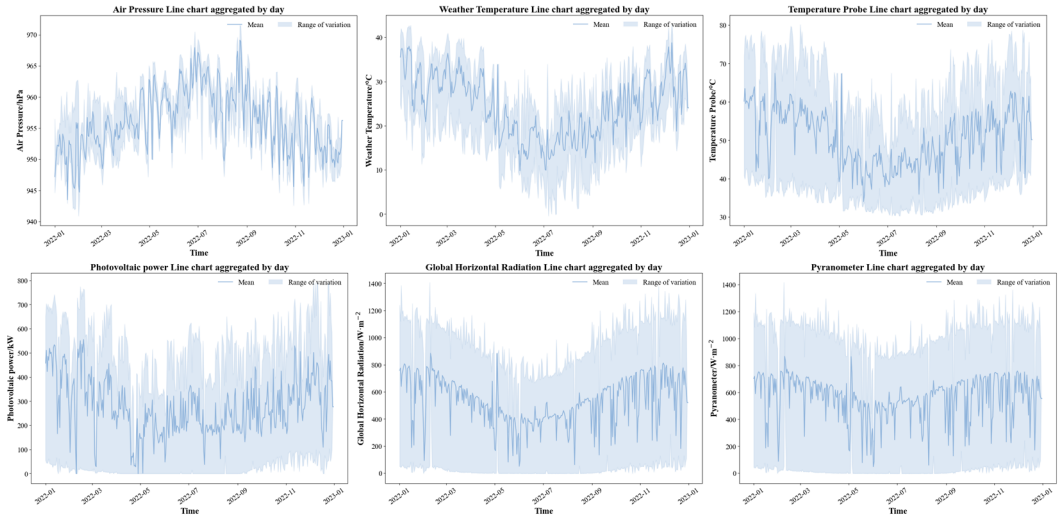

Figure 3. Trend chart of mean and extreme values of each variable

4.2. Prediction result analysis

4.2.1. Analysis of parameter optimization results of DBO-LSTM model

To ascertain the efficacy of the LSTM model optimized by the DBO algorithm, this study selected the traditional LSTM model and the DBO-LSTM model for a comparative assessment. In order to avoid experimental contingency, photovoltaic power generation data in February, May, August and November were randomly selected for analysis and modeling. Parameters of the DBO optimization algorithm were set as shown in Table 3. By analyzing the change of fitness curve and parameter optimization process, it is found that in the process of DBO parameter learning, the data in February achieved the best performance in the 4th iteration, the data in May in the 9th iteration, the data in August in the 7th iteration and the data in November in the 3rd iteration. The visual results of hyperparameter optimization process for each month are shown in Figure 4.

Table 3. Parameter setting of DBO

Parameter	Value
pop population	5
Maximum iterations	10
Ratio of rolling dung beetles	0.2
Proportion of laying beetles	0.4
Percentage of dung beetles stealing	0.2
The proportion of dung beetles	0.2
Learning rate optimization range	[0.001,10]
Neuron number optimization range	[10,150]

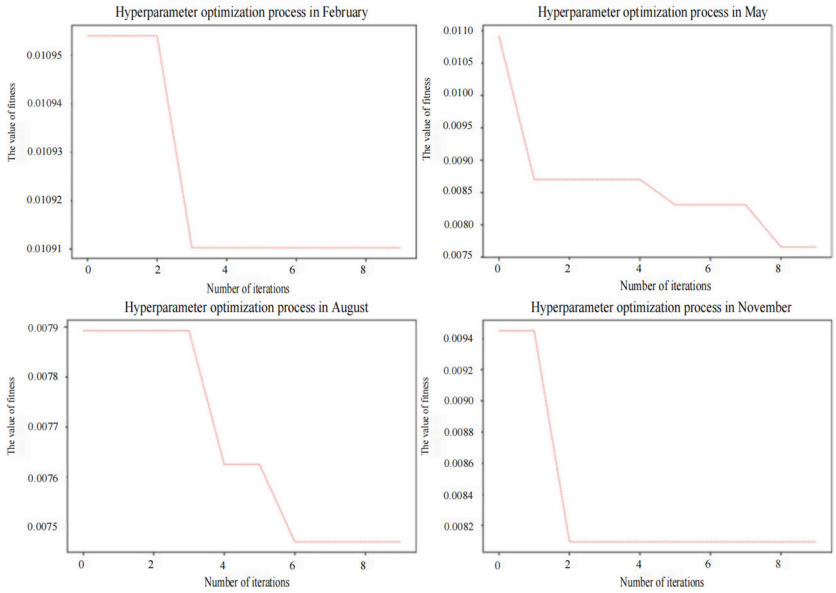

Figure 4. Hyperparameter optimization process for each month

4.2.2. Analysis of prediction results of DBO-LSTM-EBRB based on ensemble learning

This study on January 1, 2022 - December 31, photovoltaic power maximum and minimum value of load-point, within the scope of the selection with interval 6 numerical as the initial value of EBRB model results property, for [0,158.33,316.67,475.00,633.33,791.66]. At the same time, the utility value of the initial premise property is set in the same way and the weight of the premise property is set to equal weight. Set the time step to 8, the number of neurons to 50, the optimizer to the Adam optimizer, the training rounds to 10, and the batch size to 32.

The weights of EBRB model and DBO-LSTM model are set to 0.4 and 0.6 based on multiple experiments, and new fitting values are generated by ensemble learning. The parameter optimization results of EBRB model, LSTM model, DBO-LSTM model and DBO-LSTM-EBRB model were utilized to forcast the PV power generation data in 4 months respectively. The predicted results are displayed in Figure 5. The results show that the DBO-LSTM-EBRB model has the most optimal fit and performance among multiple models, indicating that the DBO-LSTM-EBRB model based on ensemble learning can effectively make up for the weekness of each single model and precision and reliability of the predictive model.

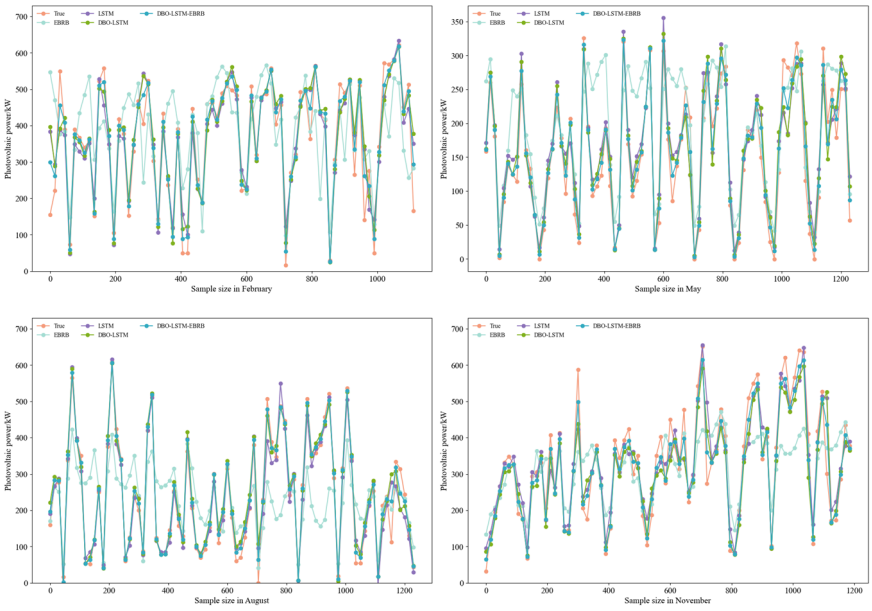

Figure 5. DBO-LSTM-EBRB Model and Single Model Monthly Forecast Results

At the same time, it is observed that the DBO-LSTM-EBRB model has excellent performance in the three key indicators, which is significantly improved compared with the traditional LSTM model, EBRB model and DBO-LSTM model. The R^2 value is 0.950-0.978, and the prediction effect is excellent. This shows that the DBO-LSTM-EBRB model has stronger ability in parameter tuning and model optimization after integrated learning of DBO-LSTM model and EBRB model.

According to the observation and analysis, photovoltaic power generation began to decline significantly from May 2022, and for a long time at a low level, indicating that due to unfavorable weather conditions in autumn and winter, and during this period of time, mostly cloudy and rainy weather, reducing the direct exposure of solar radiation, low temperature will affect the working temperature of photovoltaic panels, and then influence the power generation efficiency of photovoltaic panels.

Table 34. Comparison of prediction results between DBO-LSTM-EBRB model and single model

Month	Index	EBRB	LSTM	DBO-LSTM	DBO-LSTM-EBRB
February	R^2	0.380	0.775	0.869	0.953
	MAE	94.834	54.814	39.767	23.860
	RMSE	126.556	76.179	58.199	34.919
May	R^2	0.503	0.850	0.866	0.950
	MAE	52.385	26.444	25.463	15.493
	RMSE	68.277	37.435	35.497	21.634
August	R^2	0.343	0.923	0.939	0.978
	MAE	97.120	28.789	25.476	15.286
	RMSE	122.176	41.805	37.269	22.361
November	R^2	0.562	0.895	0.898	0.962
	MAE	77.129	36.346	35.559	21.807
	RMSE	100.232	48.954	48.273	29.372

5. Conclusion

(1) The photovoltaic power generation system is susceptible to seasonal fluctuations in the local climate. Since May 2022, there has been a notable decline in photovoltaic power generation, which has remained at a low level for an extended period. This decline can be attributed to unfavorable weather conditions during the autumn and winter seasons, which have resulted in reduced direct solar irradiation and elevated temperatures, both of which have negatively impacted the performance of photovoltaic panels. Consequently, this affects the power generation efficiency of the PV panels.

(2) Global radiation and temperature are significant factors affecting the power of photovoltaic power generation. As the global radiation and temperature gradually decreased from January to July, the photovoltaic power generation showed a downward trend, and an upward trend from July to December.

(3) Due to the comprehensive effect of multiple factors, the forecast analysis of 2022 photovoltaic power generation data shows that the integrated The DBO-LSTM-EBRB model exhibits superior predictive performance relative to the traditional LSTM and EBRB models, as evidenced by the R2 values of 0.950-0.978. The predictive efficacy of the DBO-LSTM-EBRB model is highly commendable.

References

[1] Ding Yiting. Renewable energy development got off to a good start this year [N]. People's Daily, 2023-05-05 (002) DOI: 10.28655 / n.c. Nki NRMRB. 2023.004612.

[2] zhi-peng wang. On the basis of renewable energy power system load frequency control [D]. Linyi university, 2023. The DOI: 10.44252 /, dc nki. Glydx. 2023.000151.

[3] LI Jianning, Liu Qihui, Li Ying et al. Research and development of key technologies of photovoltaic power generation [J]. Electrical Applications, 2012, 31(5): 70-76.

[4] Han Chunlei. Discussion on the status quo and development of solar photovoltaic power generation technology [J]. Light Sources and Lighting, 2022, (03): 95-97.

[5] Zhang Jingzhu, He Yuxin. Application of grid-connected photovoltaic technology in power system [J]. Light Sources and Lighting,2023,(10):139-141.

[6] Zhou Yucai, Xiao Tian, Xie Qiyue, et al. Short-term prediction of HPO-BILSTM photovoltaic power based on clustering [J]. Acta Solar Energy Sinica,2024,45(04):512-518.DOI:10.19912/ J.0254-0096.tynxB.2023-0402.

[7] Lin Z. Optimization design and economic analysis of energy storage system for wind power generation and photovoltaic power generation [J]. Electrical Technology and Economy,2024,(04):221-223.

[8] Ma Jun, Jiang Rui, Ding Qian, et al. Short-term photovoltaic power prediction based on multiverse optimization support vector machine [J]. Thermal Power Generation, 2020, 49(4): 87-92.

[9] ZHU H, LI X, SUN Q, et al. A Power Prediction Method for Photovoltaic Power Plant Based on Wavelet Decomposition and Artificial Neural Networks [J]. Energies, 2016, 9(1): 11.

[10] Zhang Jin, Liu Yun, Peng Shurong. Prediction of GRU-A photovoltaic power generation based on feature mining [J]. Laboratory Research and Exploration, 2020, 39(5): 25-30+49.

[11] Wang Xiaoxia, Yu Min, Huo Zejian et al. Short-term power prediction of distributed photovoltaic power plant Cluster based on Near neighbor Propagation clustering and LSTNet [J]. Automation of Electric power Systems, 2022, 47(6): 133-141.

[12] NIU D, WANG K, SUN L, et al. Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study [J]. Applied Soft Computing, 2020, 93: 106389.

[13] Lu Jing, Qi Haidi, Li Baode. Pirate attacks based on extended belief rule base risk prediction [J]. Journal of transportation systems engineering and information technology, 2022, 22 (3) : 247-254 + 266. DOI: 10.16097 / j.carol carroll nki. 1009-6744.2022.03.028.