Current Science
CS 6 (2), 6180-6196 (2026)
Received 27 February 2026 | **Accepted 31 March** 2026 | **Online Available 06 April** 2026
https://doi.org/10.52845/CS/2026-6-2-11

Current Science

Current Science

OPEN ACCESS

ISSN (O) 2795-8639

Original Article

Research on Short-Term Photovoltaic Power Generation Forecasting Based on GA-BiLSTM-Attention

Kai Xiao[1], Xue Liu[1], Cheng Dai[1], Zhijiang Wen[1], Jinwei Ding[2*]

[1]Guangxi Transport Investment Technology Co., Ltd. 369 Minzu Avenue, West End of Liulu Bridge, Nanning,Guangxi Zhuang Autonomous Region, China

[2]China Railway Chengdu Research Institute, Science and Technology Co., Ltd. Chengdu, SiChuan, China

***Corresponding Author: Jinwei Ding**

Abstract:
Photovoltaic (PV) power generation exhibits randomness and volatility, making high-accuracy day-ahead forecasting crucial for grid stability. This paper proposes a hybrid model, GA-BiLSTM-Attention, combining multi-scale Gaussian filtering decomposition with genetic algorithm optimisation to address issues such as prediction delay and model generalisation. Core features, irradiance and backplate temperature, are selected based on the Spearman correlation coefficient. Multi-scale Gaussian filtering decomposes historical power data into four intrinsic mode functions (IMFs), with delay decomposition to prevent data leakage. A BiLSTM-Attention architecture captures both forward and backward temporal dependencies, while attention focuses on key features. The genetic algorithm optimises hyperparameters like hidden layer dimensions, learning rate, and dropout rate. Validation using data from a 9.2 kWp PV platform in Xichang City shows that the proposed model achieves RMSE, MAE, and MAPE of 0.287 kW, 0.217 kW, and 13.81%, respectively, outperforming baseline LSTM and other comparative models. This model provides a high-accuracy, low-complexity solution for day-ahead PV power forecasting.
Keywords: Photovoltaic Power Forecasting; Bidirectional Long Short-Term Memory Network; Attention Mechanism; Genetic Algorithm; Multi-Scale Gaussian Filtering

Introduction

In the global context of addressing climate change and pursuing sustainable development, photovoltaic (PV) power generation technology, with its low-carbon and widely distributed advantages, has become one of the mainstream energy sources for global electricity consumption (Szabo et al., 2024). PV systems primarily rely on the photovoltaic effect to convert solar energy into electrical power. However, their output is directly driven by solar irradiance, cloud cover, and complex meteorological fluctuations, exhibiting strong randomness, volatility, and intermittency (Gu et al., 2025). Studies have shown that due to cloud path movement and the uncertainty of local atmospheric layers, PV output often demonstrates highly non-stationary fluctuation characteristics (Gu et al., 2025; Yuan et al., 2025).

The volatility of PV output presents significant challenges to real-time power system scheduling, frequency regulation, and reserve capacity planning (Szabo et al., 2024; Mohanasundaram & Rangaswamy, 2025). When a high proportion of PV power is integrated into the grid, insufficient prediction accuracy can easily lead to grid frequency instability, voltage violations, or even severe curtailment of PV power (Mohanasundaram & Rangaswamy, 2025).

Therefore, accurate PV power forecasting has become a crucial foundation for ensuring the safe operation of smart grids (Liu et al., 2025; Hai et al., 2025). Advanced forecasting techniques, which capture the power mapping relationship under complex meteorological conditions, not only effectively reduce spinning reserve capacity but also enhance the grid's capacity for accommodating clean energy, offering significant economic and environmental value (Hai et al., 2025).

In recent years, research on photovoltaic (PV) power forecasting has shifted from traditional statistical models to deep learning ensemble models, gradually incorporating signal decomposition, attention mechanisms, and intelligent optimisation algorithms. To address the non-stationary nature of PV power series, signal decomposition techniques have been widely applied to extract multi-scale temporal features. Gaussian filtering, as a classic smoothing method, effectively suppresses high-frequency noise while retaining trend components. Syed et al. (2021) employed moving regression filtering combined with Gaussian filtering for smoothing PV power, demonstrating the advantages of Gaussian filtering in noise suppression. Wang et al. (2025) proposed a solar radiation forecasting method based on multi-scale Gaussian data augmentation. By using variational mode decomposition (VMD), the original series is decomposed into different frequency components, and multi-scale Gaussian augmentation is applied to the training set, significantly improving multi-step prediction accuracy. Sun et al. (2025) proposed a multi-channel LSTM forecasting method based on multi-scale PV power fluctuation characteristics, proving the effectiveness of multi-scale decomposition in capturing multi-frequency features of the power series. In addition, wavelet decomposition (WPD) (Liu et al., 2021) and empirical wavelet transform (EWT) (Jaihuni et al., 2021) have also been shown to effectively reduce the non-stationarity of the series.

In terms of temporal feature extraction, Long Short-Term Memory (LSTM) networks, with their gating mechanisms, are capable of effectively capturing long-term dependencies in sequential data (Liu et al., 2021; Jaihuni et al., 2021). To further capture bidirectional temporal dependencies, Bidirectional LSTM (BiLSTM) has been introduced into the PV forecasting domain. Geng et al. (2023) combined BiLSTM with time encoding (Time2Vec) and weighted deep convolutional neural networks (WDCNN) for hybrid wind and solar power forecasting, demonstrating that the bidirectional architecture makes better use of both forward and backward meteorological information. Bashir et al. (2025) proposed a CNN-ABiLSTM hybrid model, which showed excellent performance in weekly and monthly forecasting tasks. Wang et al. (2023) combined Gradient Boosting Decision Trees (GBDT) with BiLSTM to enhance the non-linear mapping ability in daily forecasting tasks.

To address the issue of information loss in long sequence training, the Attention Mechanism has been introduced into forecasting frameworks to dynamically allocate key temporal weights. Hu et al. (2024) proposed an LSTM model with a self-attention mechanism, which efficiently integrates historical power and real-time weather forecast data. Zhou et al. (2025) developed a photovoltaic power forecasting framework based on attention mechanisms and parallel prediction architecture, achieving a breakthrough in ultra-short-term forecasting accuracy. Chen et al. (2025) proposed a CNN-BiLSTM-Attention model optimised by Bayesian methods, which achieves high-accuracy interval forecasting by modelling conditional error dependencies. Rai et al. (2023) introduced the Differentiated Attention Network (DAN), demonstrating that multi-dimensional feature allocation enhances the model's robustness under abrupt weather changes. In addition, Wan et al. (2023) and Pan et al. (2023) applied the CNN-

LSTM-Attention architecture to combined heat and power load forecasting and oil well production prediction, respectively, validating the architecture's versatility in time series forecasting.

Deep learning models are highly sensitive to hyperparameter configurations, and metaheuristic algorithms are widely used for parameter optimisation. Genetic Algorithm (GA), as a classic global optimisation method, has demonstrated its strong capability in deep learning hyperparameter search. Balaha et al. (2021) applied GA to optimise deep learning model hyperparameters, successfully applying it to COVID-19 image segmentation and recognition tasks. Suhail et al. (2023) used evolutionary genetic algorithms to optimise YOLOv5 model hyperparameters, demonstrating that GA can effectively avoid the local optimum trap of grid search. In the field of photovoltaic forecasting, Yu et al. (2023) used the Whale Optimisation Algorithm (WOA) to optimise BiLSTM-Attention model parameters, combining weather classification and dual-layer decomposition strategies to effectively improve forecasting accuracy. Peng et al. (2024) applied Particle Swarm Optimisation (PSO) combined with BiLSTM for wind and solar power joint forecasting, achieving self-adaptive parameter optimisation.

In summary, existing research indicates that: BiLSTM, with its bidirectional temporal modelling capabilities, can fully capture the forward and backward dependencies of the power series; the attention mechanism, through dynamic weight allocation, can effectively focus on key temporal information; genetic algorithms show their advantage in avoiding local optima in global hyperparameter searches; and multi-scale Gaussian filtering decomposition can effectively extract multi-frequency features of the power series while suppressing high-frequency noise.

Research on Forecasting Models

This paper proposes a hybrid forecasting model, GA-BiLSTM-Attention, which integrates multi-scale Gaussian filtering decomposition and genetic algorithm optimisation. The model extracts multi-frequency features from the power series using multi-scale Gaussian filtering decomposition, captures bidirectional temporal dependencies through BiLSTM, dynamically focuses on key periods using an additive attention mechanism, and automatically searches for the optimal hyperparameter configuration through genetic algorithms. The goal is to achieve high-accuracy and robust day-ahead photovoltaic power forecasting, providing reliable technical support for grid scheduling and energy management.

The forecasting process for day-ahead photovoltaic power generation based on GA-BiLSTM-Attention is shown in Figure 1. First, the data is pre-processed, and parameters that have a significant correlation with photovoltaic power generation are selected as feature parameters for training the model. Next, based on Gaussian filtering, the historical power generation data is decomposed in the time domain to obtain different frequency Intrinsic Mode Functions (IMFs), capturing both the long-term variation and local fluctuation characteristics of the power. Finally, the feature parameters are reconstructed and input into the GA-BiLSTM-Attention forecasting model, which predicts the next day's photovoltaic power generation.

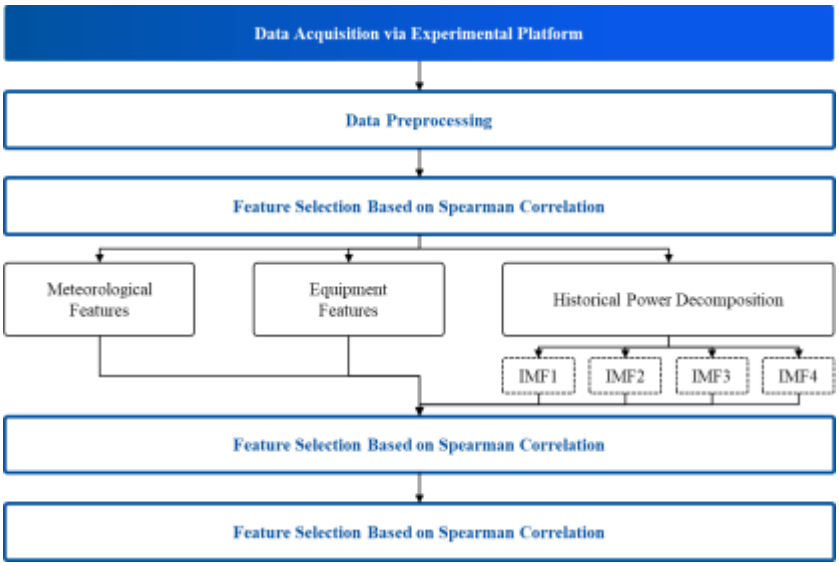

Figure 1 GA-LSTM Short-Term Power Generation Forecasting Methodology

Processing of Forecasting Feature Parameters

Feature Parameter Correlation Analysis

The Spearman correlation coefficient (ρ_s) is a statistical measure used to assess the degree of correlation between feature vectors and target vectors. Based on this metric, the correlation between various parameters and photovoltaic power generation is calculated, thereby selecting the most relevant features for predicting the next day's PV power generation and improving forecast accuracy. The detailed architecture of the model can be found in Appednix A-1.

Power Feature Extraction

The power series of photovoltaic (PV) generation exhibits significant non-stationarity and multi-scale fluctuation characteristics. To extract its intrinsic modal features, this paper employs a multi-scale Gaussian filter decomposition method. The power series is smoothed using Gaussian filters with varying window widths, enabling the extraction of fluctuation components across a range of frequencies. This approach decomposes the original power series into several sub-series (IMFs) corresponding to different frequency scales. In comparison to traditional EMD/VMD methods, Gaussian filter decomposition offers superior computational efficiency and effectively mitigates the issue of mode mixing. Let the original power generation time series be $P(t)$, where $t = 1,2,3,\ldots,N$ represents the sampling time. The initial residual sequence is $r_t = P(t)$. The detailed architecture of the model can be found in Appednix A-2.

To prevent data leakage and meet the requirements of short-term forecasting, a delayed decomposition mechanism is employed: the power data of day T is decomposed to obtain the IMF components, which are then used as input features for day $T + 1$. This strategy ensures that only historical data is used during forecasting by introducing a time shift.

Building the Prediction Model

Bidirectional Long Short-Term Memory (Bi-LSTM)

LSTM (Long Short-Term Memory) is a key deep learning algorithm in time-series forecasting. The input gate determines which data should be stored in the memory cell, and the forget gate filters out useful data. The processed data is then passed through the output gate to produce the final prediction. Its structure is shown in Figure 2. Here, $x_t$represents the input value at time t, $h_t$denotes the hidden layer output at time t, and $C_t$is the memory cell value at time t. The symbol $\otimes$represents element-wise multiplication, and $\oplus$represents element-wise addition. The detailed

architecture of the model can be found in Appednix A-3.

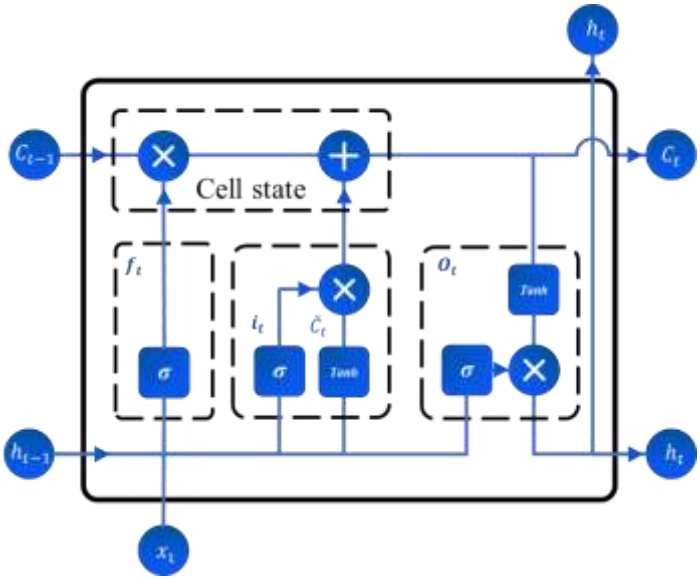

Figure 2 LSTM Network structure

The BiLSTM network, built upon a single LSTM network, incorporates a backward LSTM network. This allows it to simultaneously utilise both past data (forward information) and upcoming data (backward information), enabling a more comprehensive understanding of the trends and changes in the time series. As a result, it can more flexibly capture the temporal characteristics of photovoltaic power generation. The network structure is illustrated in Figure 3.

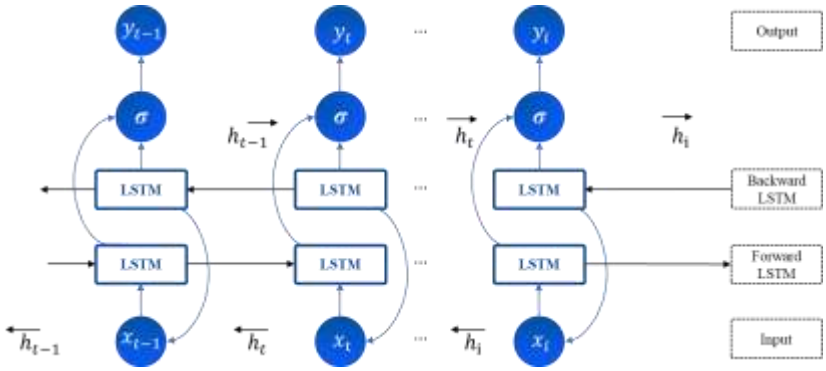

Figure 3 The structure of a BiLSTM network

Attention Mechanism

The introduction of the attention mechanism allows the model to automatically learn and focus on key temporal features: it assigns higher weights to the inflection points of the power curve, where irradiance changes sharply (such as during cloud cover or at sunrise and sunset), thus enhancing the model's ability to capture sudden changes. It also uses the IMF high-frequency components to identify local anomalous fluctuations and suppress noise interference. Furthermore, the attention distribution is dynamically adjusted under different meteorological conditions, significantly improving the model's robustness in complex scenarios. The detailed architecture of the model can be found in Appednix A-4.

Hyperparameter Optimization Based on Genetic Algorithm

To overcome the limitations of manual hyperparameter tuning and improve the prediction accuracy of the model, this paper introduces a Genetic Algorithm (GA) to automatically search for the optimal hyperparameter combination for the BiLSTM-Attention network. GA simulates the natural selection mechanism to efficiently perform global optimization of hyperparameters, such as hidden layer dimensions, learning rate, dropout rate, and batch size, thus avoiding getting trapped in local optima. The optimization process is illustrated in Figure 4.

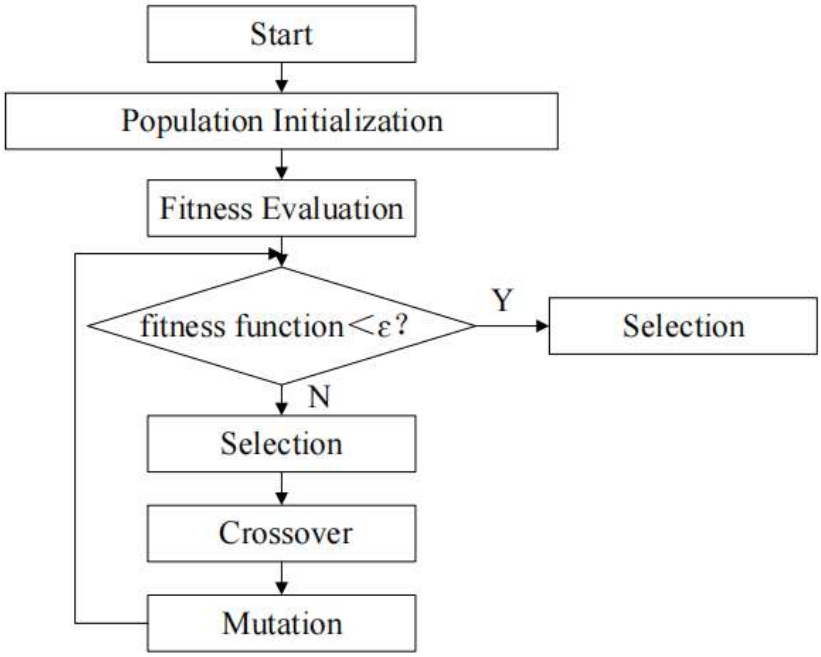

Figure 4 GA Optimization Process

GA-BiLSTM-Attention Model Architecture

The GA-BiLSTM-Attention prediction model proposed in this paper mainly consists of two layers of bidirectional LSTM, one additive attention layer, and two fully connected decoders. The hidden layers are composed of BiLSTM with 64 channels and an additive attention mechanism that incorporates Dropout regularization (0.1–0.5). The activation function used is Tanh. The overall architecture is shown in Figure 5.

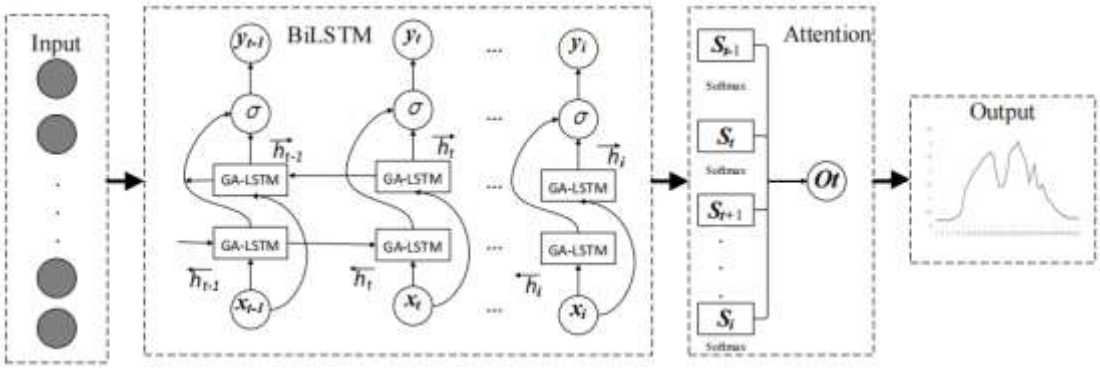

Figure 5 BiLSTM-Attention

Experiments and Analysis

Experimental Platform Setup

To obtain real and effective photovoltaic system power generation data and meteorological data, a photovoltaic system with a total installed capacity of 9.2 kWp was established in a certain area of the Chengdu-Xichang Expressway, which is under construction. A micro meteorological station was also configured. This experimental platform is a self-consumption photovoltaic system, consisting of four photovoltaic arrays, each connected to a separate inverter. These inverters are linked to the AC bus via a distribution box to supply power to the monitoring, lighting loads, and the meteorological data collection system. The system architecture of the experimental platform is shown in Figure 6.

The real-time operational data of the experimental platform is transmitted via a wireless network from the inverter communication system to a local smart gateway, and then sent to the cloud data platform, providing a reliable and authentic data foundation for the photovoltaic power generation forecasting model.

The system collects data at a frequency of once every 15 minutes, from July 2, 2024, to March 10, 2025, during the period from 6:37 AM to 7:37 PM each day. This data includes power generation, environmental meteorological data, and parameters such as the backsheet temperature of the photovoltaic system.

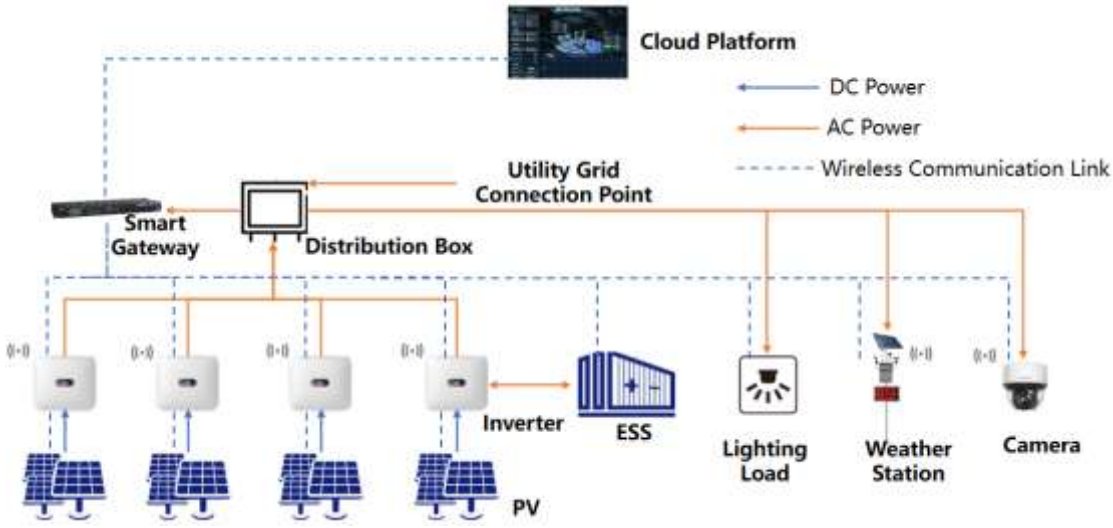

Figure 6 Experimental Platform System Architecture

Feature Parameter Processing

Feature Parameter Selection

Based on the Spearman correlation coefficient, the relationship between the parameters of the photovoltaic experimental platform and the photovoltaic power generation under different meteorological conditions was analyzed. The coefficient calculation results under cloudy, rainy, and sunny conditions are shown in

Table **1**. From the analysis, it is evident that the correlation between irradiance, backsheet temperature, and power generation exceeds 0.5 in different weather conditions, indicating a significant relationship.

To avoid interference from highly correlated parameters that could affect the training accuracy of the prediction model, this study selects irradiance and backsheet temperature as the meteorological and device feature parameters, respectively. To meet the practical engineering

requirements, the meteorological parameters of day T+1 are used as input features for the model on day T+1.

Table 1 Correlation Evaluation Results

Parameter	Rainy (ρ)	Cloudy (ρ)	Sunny (ρ)
Irradiance	0.77	0.77	0.81
Backsheet_Temp	0.58	0.63	0.70
Humidity	-0.47	-0.46	-0.49
Wind_Speed	0.39	0.42	0.48
Temperature	0.29	0.33	0.43
Pressure	-0.04	0.03	0.06

Power Generation Decomposition

To extract multi-scale temporal features from the photovoltaic power generation series, this study employs a multi-scale Gaussian filtering decomposition method. The original power series is decomposed into four intrinsic mode functions (IMF) at different frequency scales. To avoid data leakage and meet the forecast requirements for the next day, a one-day delay strategy is used. Specifically, the power data from day T is decomposed to obtain the IMF components, which are then used as input features for the model on day T+1.

1. **IMF1** has a window width of 48, capturing the baseline changes in power.
2. **IMF2** has a window width of 12, capturing hourly-level fluctuations.
3. **IMF3** has a window width of 3, extracting transient responses caused by irradiance changes.
4. **IMF4** has a window width of 1, preserving system noise and micro-level disturbances.

As an example, the decomposition result for the power generation on December 25, 2024, is shown in Figure 7.

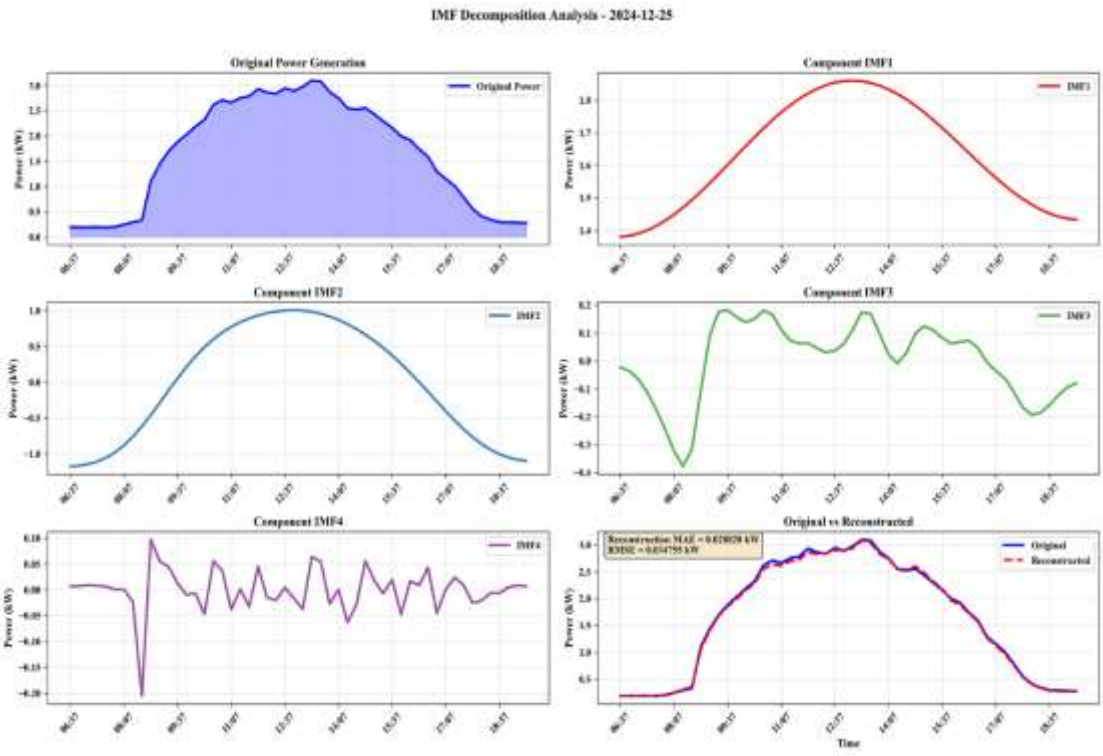

Figure 7 Power Generation Decomposition Results

By summing the four IMF components to reconstruct the original sequence, the reconstruction error was calculated with the following results: MAE (Mean Absolute Error) = 0.029 kW, RMSE (Root Mean Square Error) = 0.035 kW, and the maximum single error was only 0.075 kW, with a relative error of 1.75% (well below the 5% engineering threshold). These results validate the lossless nature and numerical stability of the decomposition algorithm, providing high-quality, low-redundancy feature inputs for the subsequent BiLSTM-Attention model.

Prediction Model Parameter Settings

The GA-BiLSTM-Attention model constructed in this study has a BiLSTM-Attention network architecture at the base level, with the top layer utilizing a Genetic Algorithm (GA) to automatically search for the optimal hyperparameter configuration. GA simulates the natural selection process, iteratively performing selection, crossover, and mutation operations within a population of candidate solutions to automatically determine the key hyperparameters of the BiLSTM-Attention model.

Specifically, GA optimizes the following hyperparameters:

1. **Hidden Layer Dimensions** (32–128): To balance model capacity with generalization ability.
2. **Learning Rate** (10^{-4}–10^{-2}): To accelerate convergence and avoid oscillations.
3. **Dropout Rate** (0.1–0.5): To dynamically control regularization strength.
4. **Batch Size** (4–16): To adapt to different data scales.
5. **Weight Decay** (10^{-5}–10^{-3}): To constrain the weight space.

This optimization process uses the RMSE of the validation set as the fitness function over 10 generations of evolution, significantly improving the model's robustness and accuracy in forecasting under sunny, cloudy, and rainy conditions. The core parameter settings of the prediction model are shown in Table 2.

Table 2 Core Architecture Parameters and GA Optimization Configuration

Parameter Category	Parameter Name	Value/Search Range
BiLSTM-Attention Architecture	Number of Layers	2-layer BiLSTM
	Hidden Dimension	96 (GA-optimized)
	Attention Mechanism	Additive Attention
	Dropout Rate	0.157 (GA-optimized)
	Input Feature Dimension	9 dimensions
Genetic Algorithm Optimization	Optimized Parameters	5 hyperparameters
	Evolution Generations	10 generations
	Population Size	12 individuals
	Fitness Function	Validation RMSE
Training Configuration	Learning Rate	0.00147 (GA-optimized)
	Weight Decay	0.000592 (GA-optimized)
	Batch Size	4 (GA-optimized)

Evaluation Metrics

To scientifically evaluate the performance of the model, three evaluation metrics — the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) — are used to analyze the model's prediction results. The detailed architecture of the model can be found in Appednix A-5.

Evaluation and Analysis

Based on the GA-BiLSTM-Attention prediction model, 70% of the experimental platform data was used as the training set, 15% as the validation set, and 15% as the test set. The performance of the prediction model is shown in Figure 8.

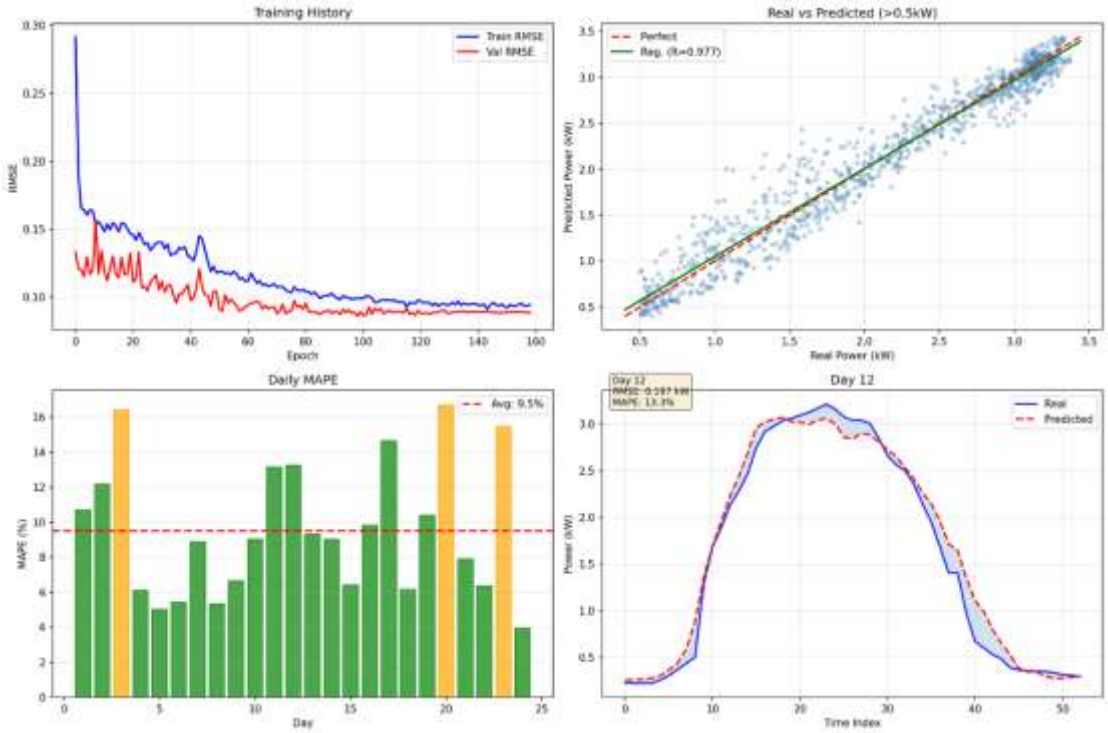

Figure 8 GA-BiLSTM-Attention Model Performance

As observed from the figure, the GA-BiLSTM-Attention model demonstrates excellent prediction performance on the test set. During the model training process, the RMSE of both the training and validation sets converged quickly and remained stable, with early stopping achieved at the 159th epoch, indicating no overfitting. The predicted power closely aligned with the actual power on either side of the ideal regression line, with a correlation coefficient R of 0.977, showcasing the model's strong fitting ability. For example, the prediction for the 12th day showed that the predicted power curve nearly overlapped with the actual curve, confirming the model's high-fidelity restoration capability for key features such as power ramp-up and peak power. The MAPE was only 13.3%.

Overall, the model achieved an accuracy level of RMSE = 0.1928 kW, MAPE = 9.44%, and $R^2 = 0.9535$for valid data points with power thresholds above 0.5 kW. The average peak prediction error was only 0.121 kW (with a relative error of 4.1%). Particularly, for the power range of 1–3 kW, the model exhibited a MAPE of 9.5% and an $R^2$of 0.932, fully validating the effectiveness and reliability of the proposed method for day-ahead photovoltaic power prediction.

To further validate the scientific and rational nature of the GA-BiLSTM-Attention prediction model, several representative models were selected for comparison and analysis. These models include BiLSTM, LSTM-Attention, CNN, CNN-Attention, and Transformer. The prediction results for the test set of each model are shown in Figure 9, and the error analysis results are presented in Table 3.

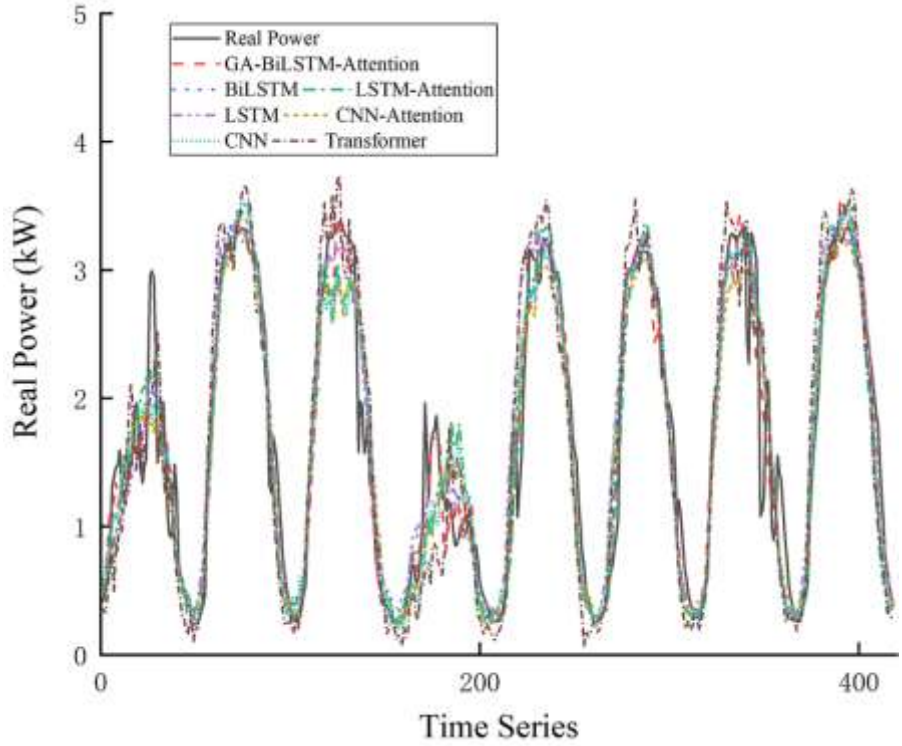

Figure 9 Comparison of Prediction Results for Different Algorithms

Table 3 Error Comparison of Different Algorithms

Models	X_{RMSE}/kW	X_{MAE}/kW	X_{MAPE}/%
GA-BiLSTM-Attention-Complete	0.287	0.217	13.81
BiLSTM	0.306	0.227	15.35
LSTM-Attention	0.34	0.259	16.98
LSTM	0.346	0.251	17.25
CNN-Attention	0.368	0.284	19.36
CNN	0.374	0.288	19.95
Transformer	0.477	0.383	24.62

The GA-BiLSTM-Attention model proposed in this paper demonstrates the best prediction performance on the test set. As shown in Figure 8 , the predicted power curve from this model closely matches the actual power curve, especially exhibiting superior tracking ability in the power peak region compared to the other models. In contrast, the BiLSTM, LSTM-Attention, and traditional LSTM models all exhibit varying degrees of delay and underestimation, while the CNN-based models and Transformer show weaker response capabilities during power spikes.

The quantitative evaluation results in Table 3 further validate these observations. The proposed model leads in all three core metrics — RMSE, MAE, and MAPE — with values of 0.287 kW, 0.217 kW, and 13.81%, respectively. These results represent reductions of 17.1%, 13.5%, and 19.9% compared to the baseline LSTM model. Notably, introducing only BiLSTM or the Attention mechanism as standalone improvements (BiLSTM and LSTM-Attention) reduced the MAPE to 15.35% and 16.98%, but these values are still significantly higher than the overall optimized performance of the proposed model.

From the architecture comparison, CNN-based models are limited by their local receptive field,

resulting in a MAPE between 19.36% and 19.95%. The Transformer model, on the other hand, suffers from limited performance due to the small dataset, with a prediction error of 24.62%. The proposed model, through GA optimization of hyperparameters, BiLSTM's bidirectional temporal modeling, and the Attention mechanism's feature focus, effectively extracts the nonlinear fluctuation characteristics of photovoltaic power. It maintains high precision across the entire global range while excelling at capturing power peaks. This demonstrates the superiority of the proposed method in day-ahead power forecasting tasks.

Conclusion and Discussion

This paper addresses the challenge of improving the accuracy of day-ahead photovoltaic power generation forecasting. The proposed approach focuses on three key aspects: optimization of hyperparameters using a genetic algorithm (GA), the design of a BiLSTM-Attention architecture, and a comparative analysis with other predictive models. The effectiveness and superiority of the proposed method are thoroughly validated, with the following key conclusions.

To tackle the complexity of hyperparameter tuning in the BiLSTM-Attention network, the inefficiency of manual adjustments, and the risk of converging to local optima, this paper introduces a genetic algorithm to automatically search for optimal values for key hyperparameters, including hidden layer dimensions, learning rate, dropout rate, batch size, and weight decay. The optimization process uses the RMSE of the validation set as the fitness function, and after 10 generations, the GA converges to an optimal configuration (hidden layer dimension = 96, learning rate = 0.00147, dropout rate = 0.157). This strategy eliminates the need for manual tuning, ensuring faster convergence and significantly improving the training efficiency and generalization robustness of the model. This approach also enables early stopping at the 159th epoch, avoiding overfitting and providing an effective paradigm for automating hyperparameter tuning in complex deep learning models.

The proposed two-layer BiLSTM architecture effectively captures bidirectional temporal dependencies in the power generation time series through both forward and backward sequence encoding. The additive attention mechanism assigns higher weights during critical moments of rapid irradiance changes, such as cloud cover, sunrise, and sunset, thus enhancing the model's ability to recognize sudden power variations. By leveraging the complementary strengths of BiLSTM and attention mechanisms, the model achieves an impressive accuracy of MAPE = 13.81% and $R^2 = 0.9535$in forecasting effective power values above 0.5 kW. This demonstrates the model's superior ability to capture key dynamic features, such as power ramp-up and peak power, in comparison to traditional sequence models.

A quantitative comparison with several baseline models — including BiLSTM, LSTM-Attention, LSTM, CNN-Attention, CNN, and Transformer — shows that the proposed GA-BiLSTM-Attention model outperforms all others in terms of RMSE (0.287 kW), MAE (0.217 kW), and MAPE (13.81%). These performance metrics are approximately 20% better than the baseline LSTM model, and 1.54 percentage points lower than the next best-performing BiLSTM model. The CNN-based models, due to their limited local receptive fields, exhibit a MAPE in the range of 19.36% to 19.95%, while the Transformer model, constrained by the small dataset, shows a higher error of 24.62%. These results further confirm the comprehensive performance advantages of the proposed method and its engineering applicability for day-ahead photovoltaic power forecasting tasks.

The results confirm that the combination of GA optimization, the BiLSTM architecture, and the

attention mechanism significantly enhances the model's ability to predict photovoltaic power generation. The GA efficiently explores the hyperparameter space, providing a robust and automated tuning approach. The BiLSTM architecture's bidirectional modeling captures the full range of temporal dependencies in the power generation data, while the attention mechanism improves the model's sensitivity to sudden changes in irradiance. As demonstrated in the comparative analysis, the proposed model's superior performance in both accuracy and generalization capability underscores its potential for real-world applications.

However, it is important to note that the model's performance could be further improved with larger datasets, particularly in regions with more complex weather patterns. Future research could explore combining additional meteorological features or experimenting with more advanced deep learning techniques, such as transformer-based architectures, to further enhance prediction accuracy. Despite these opportunities for further development, the proposed method already represents a significant advancement in photovoltaic power forecasting.

In summary, the GA-BiLSTM-Attention model provides an effective solution for day-ahead photovoltaic power forecasting, with significant improvements in both model performance and training efficiency. The model's design, based on the synergy between GA optimization, BiLSTM architecture, and attention mechanisms, ensures high-accuracy predictions and robust generalization. The quantitative comparisons confirm the proposed method's superiority over other state-of-the-art algorithms, highlighting its practical value for real-world forecasting applications.

References

1. Balaha, H. M., Balaha, M. H. & Ali, H. A. Hybrid COVID-19 segmentation and recognition framework (HMB-HCF) using deep learning and genetic algorithms. *Artificial Intelligence in Medicine* 119, 1021 56 (2021).
2. Bashir, T., Wang, H. & Tahir, M. Wind and solar power forecasting based on hybrid CNN-ABiLSTM, CNN-transformer-MLP models. *Renewable Energy* 237, 121532 (20 25).
3. Chen, Y., Wang, X. & Huang, R. Photovoltaic power interval prediction with conditional error dependency using Bayesian optimized deep learning. *Scientific Reports* 15, 1835 (2025).
4. Dash, P. K., Majumder, I. & Nayak, N. Point and interval solar power forecasting using hybrid empirical wavelet transform and robust wavelet kernel ridge regression. *Natural Resources Research* 29, 2975–2998 (2020).
5. Elsaraiti, M. & Merabet, A. Solar power forecasting using deep learning techniques. *IEEE Access* 10, 31692–31698 (2022).
6. Geng, D., Wang, B., Gao, Q. et al. A hybrid photovoltaic/wind power prediction model based on Time2Vec, WDCNN and BiLSTM. *Energy Conversion and Management* 291, 11 7327 (2023).
7. Gu, Z., Li, B., Zhang, G. et al. Optimizing photovoltaic integration in grid management via a deep learning-based scenario analysis. *Scientific Reports* 15, 1234 (2025).
8. Hai, T., Ali, A. B. M. & Agarwal, D. Predictive optimization using long short-term memory for solar PV and EV integration in relatively cold climate energy systems with a regional case study. *Scientific Reports* 15, 67 89 (2025).
9. Hu, Z., Gao, Y. & Ji, S. Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data. *Applied Energy* 359, 122709 (2024).
10. Jaihuni, M., Basak, J. K., Khan, F. et al. A

novel recurrent neural network approach in forecasting short term solar irradiance. *ISA Transactions* 121, 63–74 (2021).

11. Li, P., Zhou, K. & Lu, X. A hybrid deep learning model for short-term PV power forecasting. *Applied Energy* 259, 114216 (2020).
12. Liu, C. H., Gu, J. C. & Yang, M. T. A simplified LSTM neural networks for one day-ahead solar power forecasting. *IEEE Access* 9, 17174–17195 (2021).
13. Liu, H., Cai, C. & Li, P. Hybrid prediction method for solar photovoltaic power generation using normal cloud parrot optimization algorithm integrated with extreme learning machine. *Scientific Reports* 15, 5678 (2025).
14. Mohanasundaram, V. & Rangaswamy, B. Photovoltaic solar energy prediction using the seasonal-trend decomposition layer and ASOA optimized LSTM neural network model. *Scientific Reports* 15, 4567 (2025).
15. Pan, S., Yang, B. & Wang, S. Oil well production prediction based on CNN-LSTM model with self-attention mechanism. *Energy* 284, 128701 (2023).
16. Peng, S., Zhu, J. & Wu, T. Prediction of wind and PV power by fusing the multi-stage feature extraction and a PSO-BiLSTM model. *Energy* 298, 131345 (2024).
17. Rai, A., Shrivastava, A. & Jana, K. C. Differential attention net: Multi-directed differential attention based hybrid deep learning model for solar power forecasting. *Energy* 263, 125518 (2023).
18. Suhail, K. & Brindha, D. Microscopic urinary particle detection by different YOLOv5 models with evolutionary genetic algorithm based hyperparameter optimization. *Computers in Biology and Medicine* 169, 107895 (2023).
19. Sun, F., Li, L. & Bian, D. Photovoltaic power prediction based on multi-scale photovoltaic power fluctuation characteristics and multi-channel LSTM prediction models. *Renewable Energy* 242, 121832 (2025).
20. Syed, M. A. & Khalid, M. Moving regression filtering with battery state of charge feedback control for solar PV firming and ramp rate curtailment. *IEEE Access* 9, 13198–13211 (2021).
21. Szabo, L., Moner-Girona, M. & Jäger-Waldau, A. Impacts of large-scale deployment of vertical bifacial photovoltaics on European electricity market dynamics. *Nature Communications* 15, 6794 (2024).
22. Wan, A., Chang, Q. & Khalil, A. L. B. Short-term power load forecasting for combined heat and power using CNN-LSTM enhanced by attention mechanism. *Energy* 282, 128274 (2023).
23. Wang, S. & Ma, J. A novel GBDT-BiLSTM day-ahead PV forecasting model. *Scientific Reports* 13, 13523 (2023).
24. Wang, Z., Peng, T. & Zhang, X. A multi-step short-term solar radiation forecasting based on optimized generalized regularized extreme learning machine and multi-scale Gaussian data augmentation strategy. *Applied Energy* 377, 124499 (2025).
25. Yu, M., Niu, D. & Wang, K. Short-term photovoltaic power point-interval forecasting based on double-layer decomposition and WOA-BiLSTM-Attention and considering weather classification. *Energy* 273, 127183 (2023).
26. Yuan, Z., Xu, Y. & Xie, S. A dual-domain seasonal hybrid forecasting strategy for PV power considering dynamic uncertain fluctuations. *Scientific Reports* 15, 3456 (2025).
27. Zhou, Z., Dai, Y. & Leng, M. A photovoltaic power forecasting framework based on Attention mechanism and parallel prediction architecture. *Applied Energy* 378, 124599 (2025).

Appendix A

Appednix A-1 Feature Parameter Correlation Analysis

The expression is as follows:

$$\rho_s = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{N}(X_i - \bar{X})^2 \sum_{i=1}^{N}(Y_i - \bar{Y})^2}} \quad (1\text{-}1)$$

ρ_s is the correlation evaluation index, with a value range from -1 to 1. When the absolute value of ρ_s approaches 1, it indicates a strong correlation between the influencing parameters and power generation. Conversely, when it approaches 0, it suggests a weak correlation.

N is represents the number of samples collected;

X_i is denotes the collected data of influencing parameters, including irradiance, ambient temperature, air pressure, humidity, wind speed, and backplate temperature;

Y_i represents the collected power generation data;

$\bar{X}$ is the mean values of the influencing parameters;

$\bar{Y}$ is the mean values of the power generation.

Appednix A-2 Power Feature Extraction

The specific feature extraction process is as follows:

1) For the i-th layer of decomposition ($i = 1,2,\ldots,k-1$, where k is the preset number of decomposition layers), the Gaussian filter G_{σ_i} is applied to smooth the residual sequence, resulting in the i-th intrinsic mode function (IMF):

$$\begin{cases} \mathrm{IMF}_i(t) = (G_{\sigma_i} * r_{i-1})(t) = \sum_{\tau=-\infty}^{\infty} G_{\sigma_i}(\tau) \cdot r_{i-1}(t-\tau) \\ G_{\sigma_i}(\tau) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{\tau^2}{2\sigma_i^2}\right) \end{cases} \quad (2\text{-}1)$$

σ_i is the standard deviation, determined by the preset window parameter w_i, Assume $\sigma_i = \frac{w_i}{3}$;

$*$ denotes the discrete convolution operation.

2) The extracted mode is removed from the current residual:

$$\begin{cases} \mathrm{IMF}_i(t) = (G_{\sigma_i} * r_{i-1})(t) = \sum_{\tau=-\infty}^{\infty} G_{\sigma_i}(\tau) \cdot r_{i-1}(t-\tau) \\ G_{\sigma_i}(\tau) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{\tau^2}{2\sigma_i^2}\right) \end{cases} \quad (2\text{-}2)$$

3) After the modal components have been extracted, the residual is updated by removing the extracted mode, and the reconstructed signal is validated to ensure the accuracy and completeness of the decomposition process:

$$\begin{cases} P(t) = \sum_{i=1}^{K} \mathrm{IMF}_i(t) + \grave{o}(t) \\ \mathrm{MAE}_{\mathrm{recon}} = \frac{1}{N} \sum_{t=1}^{N} \left| P(t) - \sum_{i=1}^{K} \mathrm{IMF}_i(t) \right| \end{cases} \quad (2\text{-}3)$$

ε_t represents the reconstruction error, which denotes the residual between $P(t)$ and the reconstructed value from the IMFs.

Appednix A-3Bidirectional Long Short-Term Memory (Bi-LSTM)

1) The retention degree of the forget gate response, with its output value $f_t$ranging from [0 to 1], where a result of 0 indicates complete forgetting and a result of 1 indicates complete retention. The expression is as follows:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad (2\text{-}4)$$

In the expression, denotes the sigmoid activation function; $w_f$is the weight coefficient corresponding to the forget gate; $b_f$is the associated bias term; $t$represents the time step of the neural unit; $h_{i-1}$is the hidden state from the previous time step, and $x_t$is the input at the current time step.

2) The input gate controls the information that is input into the cell state. It calculates a value within the range of [0 to 1] based on the current input $x_t$and the previous hidden state h_{i-1}, where values closer to 1 indicate more important information. The expression is as follows:

$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \quad (2\text{-}5)$$

$$\tilde{C}_t = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (2\text{-}6)$$

In the expression:

$i_t$is the activation value of the input gate, represents the weight parameters of the input gate;

$b_i$is the corresponding bias term;

$w_c$represents the weight parameters of the candidate value;

$\tilde{C}_t$, and $b_c$are the corresponding bias terms.

3) The cell state represents the long-term memory retained by the neural network, and its expression is given by:

$$O_t = \sigma\left(W_o \cdot [h_{t-1}, x_t] + b_o\right) \quad (2\text{-}7)$$

$$h_t = O_t * \tanh(C_t) \quad (2\text{-}8)$$

In the expression:

$O_t$represents the output of the output gate at time step t.

Appednix A-4Attention Mechanism

The computational process is shown in Equations (2-9) to (2-11).

$$S_t = \tanh(W_h k_t + b_h) \quad (2\text{-}9)$$

$$a_t = \text{Softmax}(s,t) \quad (2\text{-}10)$$

$$O_t = \sum_t a_t h_t \quad (2\text{-}11)$$

In the expression:

$S_t$represents the probability distribution;

W_{ht}denotes the weight;

$k_t$refers to the hidden state value corresponding to the output $y_t$from the BiLSTM-layer;

$a_t$is the weight coefficient;

$O_t$is the output value after the attention mechanism layer has allocated the weights.

Appednix A-5Evaluation Metrics

The expressions for each evaluation metric are as follows:

$$X_{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{p}_i - p_i\right)^2} \quad (3\text{-}1)$$

$$X_{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{p}_i - p_i\right| \quad (3\text{-}2)$$

$$X_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|\hat{p}_i - p_i\right|}{p_i} \times 100\% \quad (3\text{-}3)$$

In the expression:

$y_t$represents the actual power generation of the system, and $\hat{y}_t$represents the predicted power generation.