RESEARCH ON THE APPLICATION OF A BILSTM-INFORMER HYBRID MODEL IN REAL-TIME ENHANCEMENT PREDICTION OF PHOTOVOLTAIC POWER GENERATION OUTPUT

Ping Wu
School of Intelligent Manufacturing, Sanmenxia Polytechnic, Sanmenxia, 472000, China

Abstract - Accurate prediction of photovoltaic power generation is crucial for ensuring the safe and stable operation of the power grid. To address the shortcomings of traditional Long Short-Term Memory (LSTM) networks in capturing long-term temporal correlations in power sequences, this paper designs a hybrid architecture (BiLSTM (Bidirectional Long Short-Term Memory)-Informer) that integrates a bidirectional LSM network and an Informer to improve the real-time prediction performance of power generation. The innovation of this method is reflected in three aspects: First, by integrating the bidirectional context-aware capability of BiLSTM and the long-range dependency capture mechanism of Informer, the expressive power of time series features is significantly improved; second, a dynamic data augmentation strategy is introduced, utilizing real-time noise injection and sequence reconstruction techniques to optimize the robustness and generalization ability of the training process; finally, an adaptive loss function and parameter tuning scheme are designed to address the inherent non-stationary fluctuation characteristics of photovoltaic data. Experimental verification based on actual photovoltaic power plant operation data shows that the hybrid model constructed in this paper performs excellently in key performance indicators, with a mean absolute error (MAE) of 0.52 kW, a root means square error (RMSE) of 0.68 kW, and a mean absolute percentage error (MAPE) of 0.71%, all outperforming all compared benchmark models. The results fully demonstrate the significant advantages of the proposed method in terms of prediction stability and real-time performance, providing reliable technical support for the engineering practice of intelligent energy control systems.

Keywords: PV generation forecasting, BiLSTM, Informer, Hybrid model, Real-time augmentation.

1. Introduction

With the acceleration of the global energy system's low-carbon transition, photovoltaic (PV) power generation has become a critical component of clean energy systems, with its installed capacity and grid integration scale continuously expanding [1]. However, PV output exhibits significant intermittency and stochastic fluctuations, arising from the nonlinear coupling of multi-scale environmental factors such as meteorological parameters, solar radiation cycles, and seasonal variations. These characteristics pose substantial challenges to grid dispatch, power balance, and operational stability [2]. Consequently, developing accurate and reliable PV power forecasting systems is essential for enhancing renewable energy integration capacity, improving energy management precision, and ensuring reliable power system operation [3].

Existing forecasting methods fall into two main categories: physical models and data-driven approaches. Physical models rely on numerical weather prediction and module parameters, making it difficult to accurately capture the nonlinear dynamic characteristics of actual system operation [4-5]. Statistical learning methods, such as autoregressive integrated moving average models and support vector machines, can identify certain temporal patterns but remain limited in modeling long-term dependencies and sudden fluctuations. While deep learning models like Long Short-Term Memory (LSTM) networks excel in time series modeling, they suffer from memory attenuation when processing long sequences and face increased computational burden and memory consumption for ultra-long sequences.

Transformer-based models have demonstrated remarkable performance in time series forecasting [6-13]. Notably, Informer leverages ProbSparse self-

dx.doi.org/10.17683/ijomam/issue24.2

attention and distillation encoders for efficient long-sequence modeling, effectively extracting global patterns such as periodicity [14]. However, Informer's perception of local details is relatively weak, whereas short-term fluctuations in PV power contain critical information [15].

To address these limitations, Bidirectional LSTM (BiLSTM) captures local contextual features through bidirectional learning mechanisms [16]. This study proposes a hybrid forecasting model that integrates BiLSTM with Informer: BiLSTM serves as a front-end module to extract local features [17], whose outputs are then fed into the Informer module to mine long-term dependencies. This synergistic architecture enables collaborative modeling of both short-term fluctuations and long-term patterns, effectively compensating for the shortcomings of individual models.

2. Related Work

2.1 Overview of the BiLSTM Architecture

BiLSTM is a special recurrent neural network (RNN) designed specifically for processing sequence data. It is a sequence modeling architecture improved on the standard LSTM [18]. BiLSTM is an improvement on LSTM. This architecture fuses information from previous and subsequent moments, enabling it to more accurately extract contextual associations in sequence data. Specifically, BiLSTM consists of two LSTM layers in parallel: one processes the forward part of the input sequence, calculating sequentially from the first element to the last element of the sequence; the other processes its reverse part, calculating in reverse order from the last element to the first element of the sequence. The outputs of these two layers are merged at each time step or at the end of the sequence to enhance the model's ability to capture bidirectional context. Its core improvement lies in the use of bidirectional time series processing to more comprehensively obtain contextual dependencies in the sequence [19]. This bidirectional structure enables the model to refer to both historical and future information to make inferences at the current moment [20].

The standard LSTM unit controls information transfer through a gating mechanism and mainly contains three key components [21]. The forgetting gate is responsible for screening the retained information in the previous memory state, and its calculation process is expressed as formula (1):

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

Among them, f_t represents the forgetting gate output, σ is the sigmoid, and W_f and b_f are the designated weight matrix and bias parameter.

The input gate regulates the introduction degree of new information. This structure includes two parts: information screening and candidate state generation. Its calculation process is described by formulas (2) and (3):

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

$$\tilde{C}_t = \boldsymbol{tanh}(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

The memory state at the current moment C_t is updated by weighted combination, and the update process can be expressed as formula (4):

$$C_t = f_t \square C_{t-1} + i_t \square C_t \tag{4}$$

The output gate subsequently modulates the resultant hidden state formation, which is expressed as formulas (5) and (6):

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

$$h_t = o_t \square \boldsymbol{tanh}(C_t) \tag{6}$$

BiLSTM uses two independent LSTM layers to implement bidirectional timing processing. The forward layer traverses the input in chronological order, yielding a hidden state sequence $\overrightarrow{h_t}$; The backward layer traverses the input backwards, yielding a hidden state sequence $\overleftarrow{h_t}$. A concatenation of the bidirectional hidden states yields the final output at every time step, and its calculation process is expressed by formula (7):

$$h_t = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{7}$$

The bidirectional architecture supports the model to utilize both historical and future information to obtain a more comprehensive sequence representation [22]. The hidden layer's size and the dropout rate jointly manage the model's capacity and its risk of overfitting. During actual deployment, the forward layer and backward layer can be configured symmetrically, or different hidden dimensions can be set according to specific task requirements. BiLSTM have been successfully deployed in multiple fields, including sentiment analysis, machine translation, text summarization, and named entity recognition in NLP tasks, as well as power load forecasting, stock price analysis, and meteorological data modeling in time series forecasting, etc. BiLSTM can better understand language context and nuances, thereby providing more accurate prediction or analysis results. Speech recognition: BiLSTM can effectively model the audio context before and after a given time point, improving recognition accuracy.

Time series prediction: BiLSTM can provide more accurate predictions when future context can affect past events.

2.2 Basic Knowledge of Informer Architecture

Informer, a deep learning model built on the Transformer framework, was first proposed at the AAAI 2021 conference [23-26]. The core design goal of this architecture is to balance the computational efficiency and prediction accuracy in long-sequence time series prediction tasks. Its technological breakthroughs are concentrated in three aspects: the first innovation is the use of a probabilistic sparse self-attention mechanism, which successfully breaks through the quadratic computation bottleneck of the traditional self-attention module by adaptively selecting query vectors with significant information; secondly, the self-attention distillation technology is used to achieve layer-by-layer compression of hidden layer representations, significantly reducing the memory resource consumption during model training [27]; in addition, by constructing a generative decoder structure, one-step prediction of long sequence outputs is achieved, greatly improving the model inference speed. At the empirical research level, Informer has demonstrated superior performance in multiple time series analysis scenarios such as stock price prediction and robot motion trajectory prediction, establishing its important position as an efficient time series modeling tool. The architecture consists of an EncoderStack and a Decoder. The Decoder uses GenerativeStyle prediction to generate the target sequence in a single step, avoiding the dynamic decoding required by traditional Transformers. During preprocessing, data is converted to a specific format and fed into the model. After processing by Embedding and the Encoder, the Decoder uses ProbSparseSelf-attention and FullAttention to predict the target sequence.

At the level of attention mechanism, Informer architecture substitutes the standard multi-head attention with a ProbSparse mechanism, while the computation of the traditional self-attention is mathematically represented in formula (8):

$$\boldsymbol{Attention}(Q,K,V)=\boldsymbol{Softmax}\left(\frac{QK^{\cdot}}{\sqrt{d_k}}\right)V \tag{8}$$

Where Q, K and V correspond to the query, key and value matrix respectively, and d_k represents the dimension of the key vector. Since the complexity of dot product operation squares with the sequence length, computational bottlenecks will be formed in long sequence applications [28].

According to the sparse nature of the query vector, the ProbSparse mechanism only filters key queries to participate in the operation [29]. Its sparsity measure is defined by formula (9):

$$M(q_i,K)=max_j\left(\frac{q_ik_j^{\cdot}}{\sqrt{d}}\right)-\frac{1}{L}\sum_{j=1}^{L}\frac{q_ik_j^{\cdot}}{\sqrt{d}} \tag{9}$$

Where q_i represents the i-th query vector, k_j is the j-th key vector and L is the sequence length.

2.3 Construction of BiLSTM-Informer Hybrid Model

Figure 1 shows the BiLSTM-Informer hybrid model. Accurate estimation of photovoltaic power generation plays a decisive role in maintaining the safe and stable operation of the power grid. In order to overcome the shortcomings of traditional long short-term memory networks in identifying the long-term dependence characteristics and complex time series laws of photovoltaic output, this study designed a hybrid prediction framework that integrates bidirectional long short-term memory networks and Informer architecture, aiming to simultaneously improve prediction accuracy and real-time performance. The architecture consists of three core components: the base layer captures the forward and backward correlations of time series data through bidirectional long short-term memory units to achieve effective extraction of local time series features and short-term fluctuations; the middle layer relies on the Informer framework and adopts a probabilistic sparse attention mechanism to reduce the computational load of long sequences, and uses attention distillation technology to achieve feature dimension compression, thereby enhancing the representation ability of long-term dependencies; the top layer integrates real-time data enhancement modules and adaptive optimization mechanisms to improve the generalization performance of the model through dynamic data expansion and parameter adjustment, using dynamic noise injection and sequence transformation operations to improve the diversity of training data, combined with a customized loss function designed for the nonlinear fluctuation characteristics of photovoltaic output, significantly enhancing the generalization performance and robustness of the model.

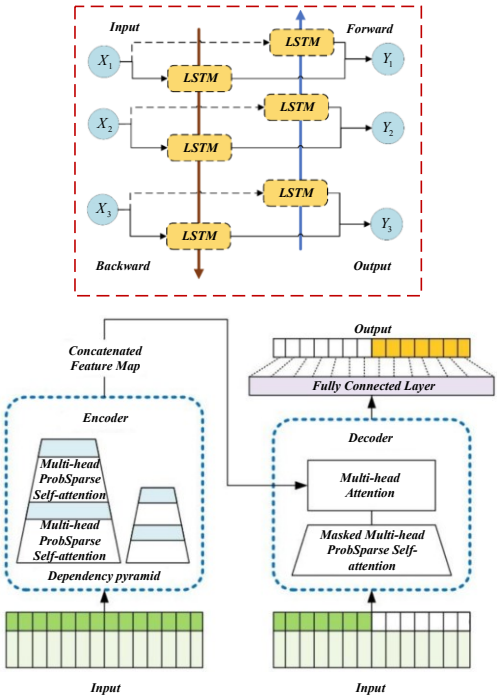

Figure 1: BiLSTM-Informer hybrid model

End-to-end integration between modules is realized through feature dimension matching. The 128-dimensional features output by the BiLSTM module are promoted to 512 dimensions through linear transformation and are used as the input source of the Informer module. The encoding vector output by the Informer module is mapped to the prediction dimension through the fully connected layer. The training process implements a phased strategy: firstly, the parameters of the Informer module are fixed, and only the BiLSTM component is trained; Then, the parameter freezing is released, and the overall model is jointly optimized. The learning rate adopts the cosine annealing scheduling scheme, and the initial value is set to 0.001.

2.4 BiLSTM-Informer Hybrid Model Training Process

In terms of model construction, the top layer integrates dynamic data augmentation and an adaptive loss function. The former expands the training samples through Gaussian noise injection (σ=0.05) and sliding window sampling. Specifically, the sliding window is set with a width of 24 time steps (corresponding to 6 hours, data resolution of 15 minutes), predicting the next 12 time steps (corresponding to 3 hours), and a step size of 6 time steps (1.5 hours). This method generates a large number of overlapping samples, enhancing the model's robustness and generalization ability to changes in the starting point. The latter integrates mean squared error and a smoothing regularization term (λ=0.01), suppressing output fluctuations while ensuring fitting accuracy and improving the physical rationality of time-series predictions.

It is important to note that to ensure the physical plausibility of the generated data, the injected noise amplitude is strictly controlled (standard deviation σ=0.05) and is applied after Z-score standardization but before being input into the model. After training, the predicted values are inversely standardized to avoid generating values that do not conform to physical laws, such as negative power. Furthermore, noise is only added to the training data as a regularization method, while the test set uses clean data to ensure the fairness of the evaluation.

The training process employs a two-stage strategy: the first stage freezes the Informer and independently trains the BiLSTM to fully extract local features; subsequently, global parameters are jointly optimized. During training, noise is dynamically injected, and cosine annealing is applied to decay the learning rate. An early stopping mechanism is also introduced (termination occurs if the validation set loss does not decrease for 10 consecutive rounds), effectively preventing overfitting and ensuring the model converges to its optimal state.

In the data preprocessing and augmentation stage, the input data includes historical photovoltaic power sequences and related meteorological variables, mainly covering time-series features such as solar irradiance and ambient temperature. First, linear interpolation is used to fill in missing data, and then the Z-score method is used to complete feature standardization.

A dynamic noise injection strategy is introduced during the training phase to add random perturbations that conform to a Gaussian distribution to the input sequence, thereby further improving the robustness and generalization ability of the model. The calculation procedure is as defined in formula (10):

$$x' = x + \grave{o}, \quad \grave{o} \sim \mathrm{N}\ (0, \sigma^2) \tag{10}$$

Where the standard deviation σ is set to 0.05. The sequence transformation adopts a sliding window mechanism to generate enhancement samples, a window width of 24 time steps and a translation interval of 6 steps are used to effectively expand the training sample size, thereby improving the generalization performance of the model.

In the first stage, the Informer parameters are frozen, and the BiLSTM part is trained separately. The BiLSTM layer captures local dependencies of the sequence through bidirectional propagation, independently calculates the hidden state variables in the forward and backward propagation processes, and then achieves feature fusion through vector concatenation. The Adam optimizer is used for training, with a learning rate set to 0.001. The regularization weight λ in the adaptive loss function is optimized through grid search and finally determined to be 0.01. An early stopping mechanism is introduced during training, which monitors the loss value on the validation set. If the loss does not decrease for 10 consecutive training epochs, training is stopped, and the model parameters are restored to the level where the validation set loss is lowest. In its gating structure, the update processes of input gate, forgetting gate and output gate are represented by formulas (11), (12) and (13) respectively:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{11}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{12}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{13}$$

Where σ is the sigmoid activation function, the hidden layer dimension is configured to 128, and the discard rate is set to 0.2 to suppress overfitting.

The second stage jointly optimizes BiLSTM and Informer module. The Informer encoder uses the ProbSparse self-attention mechanism to screen key query-key pairs through KL divergence, reducing the computational complexity to $O(L\log L)$. The attention weight is calculated in the form of formula (14):

$$\mathbf{Attention}(Q, K, V) = \mathbf{Softmax}\left(\frac{\tilde{Q}K^{\cdot}}{\sqrt{d_k}}\right)V \tag{14}$$

Where $\tilde{Q}$ is the sparse query matrix. The encoder contains a three-layer distillation structure, employing stride-2 one-dimensional convolutions, each encoder layer reduces sequence length by half with a distillation ratio of 0.7. The architecture maintains 512-dimensional representations configured with 8 attention heads.

The training process uses an adaptive loss function, which is weighted by the mean square error and the smooth regular term, and its calculation process is formula (15):

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 + \lambda\sum_{t=1}^{T-1}(\hat{y}_{t+1} - \hat{y}_t)^2 \tag{15}$$

Where the regularization weight $\lambda = 0.01$ aims to balance the prediction accuracy with the smoothness of the output sequence. The optimizer selects Adam algorithm, the learning rate is dynamically decayed from the initial value of 0.001 by cosine annealing scheduling, and the batch size is set to 64. The training period was 30 rounds, and an early stop mechanism was employed to prevent overfitting.

3. Experiment and Results Analysis

- ***Data Sets and Data Preprocessing***

The dataset contains historical power generation data and corresponding meteorological data from a real photovoltaic power plant over a continuous year, with a time resolution of 15 minutes and a total data size of approximately 1.2 GB. The power station is located in Qinghai Province, China, and has an installed capacity of 50 MWp. The data period is from January 1, 2023 to December 31, 2023, covering spring, summer, autumn, and winter to reflect the sunshine and weather patterns in different seasons. Meteorological data, including solar irradiance, ambient temperature, relative humidity, wind speed, and wind direction, are all from the meteorological monitoring station on site at the power station. Preprocessing involved missing data completion through linear interpolation and Z-score normalization. For regularization, Gaussian noise (μ=0, σ=0.05) was introduced to inputs during training.

Before training and evaluating the model, we divided the preprocessed dataset into training, validation, and test sets in chronological order. The ratio was 60%:20%:20%, meaning the first 60% of the data (approximately 7.2 months) was used for model training, the next 20% (approximately 2.4 months) was used for the validation set to adjust hyperparameters and prevent overfitting, and the last 20% (approximately 2.4 months) was used as the test set to ultimately evaluate the model's predictive performance.

This chronological division was chosen to best simulate real-world application scenarios, i.e., using historical data to predict the future.

- ***Evaluation Index***

This study uses three key evaluation parameters to systematically test the predictive performance of the model: the mean absolute error (MAE) is used to quantify the average deviation between the predicted results and the actual observed values; the root mean square error (RMSE) reflects the discrete distribution characteristics of the prediction error; and the mean absolute percentage error (MAPE) reflects the level of prediction accuracy from a relative dimension.

- ***Specific Experiments***

The data in Table 1 shows that the MAE of each model increases with the increase of prediction time. The MAE of BiLSTM-Informer at 1-hour, 3-hour, 6-hour, 12-hour and 24-hour prediction durations were 0.31, 0.52, 0.89, 1.48 and 2.35, respectively, all of which were lower than those of the comparative model. Informer was next, with a 24-hour MAE of 2.92. Transformer has the highest error under each prediction time, with a 24-hour MAE of 4.42. BiLSTM-Informer has obvious advantages in short-term prediction, and the error accumulation rate with the increase of prediction time is lower than other models.

Table 1. Comparison of MAE of each model under different prediction times (unit: kW)

Prediction duration	BiLSTM-Informer	Informer	BiLSTM	CNN-LSTM	Transformer	GRU
1 hour	0.31	0.42	0.51	0.58	0.69	0.62
3 hours	0.52	0.68	0.85	0.92	1.15	1.02
6 hours	0.89	1.15	1.42	1.58	1.92	1.75
12 hours	1.48	1.86	2.25	2.51	2.98	2.73
24 hours	2.35	2.92	3.48	3.85	4.42	4.11

According to the comparative analysis results presented in Table 2, the BiLSTM-Informer hybrid model demonstrates significant advantages in predictive performance. It not only achieves the best error index, but also achieves the highest level of determination coefficient R^2 of 0.92. This is closely followed by the independent Informer model (R^2=0.88) and the BiLSTM model (R^2=0.85).

Among the compared models, the Transformer architecture performed the weakest (R^2=0.70), while the CNN-LSTM hybrid model and the GRU model also had relatively limited predictive capabilities. These experimental data strongly demonstrate that the proposed hybrid model has a stronger explanatory power for analyzing the fluctuation characteristics of photovoltaic power generation.

Table 2. Comparison of photovoltaic power generation prediction performance of different models

Experiment No.	Model Name	Prediction error (kW)	Absolute error (kW)	RMSE (kW)	R^2
1	BiLSTM-Informer	0.52	0.68	0.71	0.92
2	BiLSTM	0.73	0.85	0.88	0.85
3	Informer	0.65	0.79	0.82	0.88
4	CNN-LSTM	0.81	0.92	0.95	0.80
5	Transformer	1.02	1.15	1.18	0.70
6	GRU	0.90	1.02	1.05	0.75

In Figure 2, the top chart shows that Informer has the lowest error across all weather conditions, with 10% error in foggy and snowy weather; BiLSTM-Informer is second, with 15% error in foggy and 12% error in snowy weather. Models like GRU and Transformer have higher errors, with GRU showing 18% error in foggy weather.

The bottom chart shows the error across different time periods. Informer outperforms other models in all time periods, with 2% error from 0-2 AM and 10% error from 12-2 PM; BiLSTM-Informer has 5% error from 0-2 AM and 15% error from 12-2 PM.

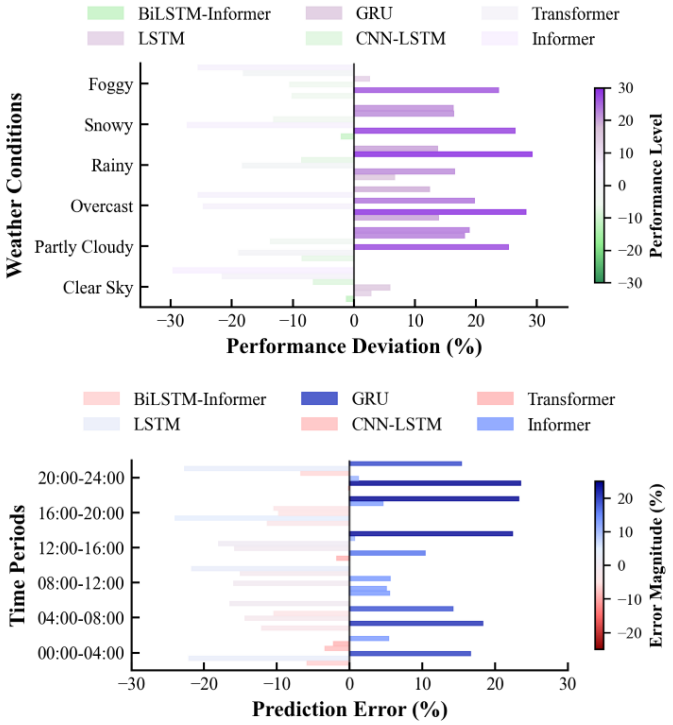

Figure 2: Comparison of prediction errors (MAPE, %) of different models under various weather conditions and time periods

As shown in Figure 3, the top-left subplot shows that the quantiles of actual and predicted values are distributed roughly along the diagonal. The top-right subplot shows that the error kernel density has a peak shape, concentrated around 0.

The bottom-left subplot shows that the MAE is highest at 0.35 in foggy weather and lowest at 0.05 in sunny weather. The bottom-right subplot shows that the MAE is highest at 0.30 from 12:00 to 14:00 and lowest at 0.04 from 0:00 to 2:00.

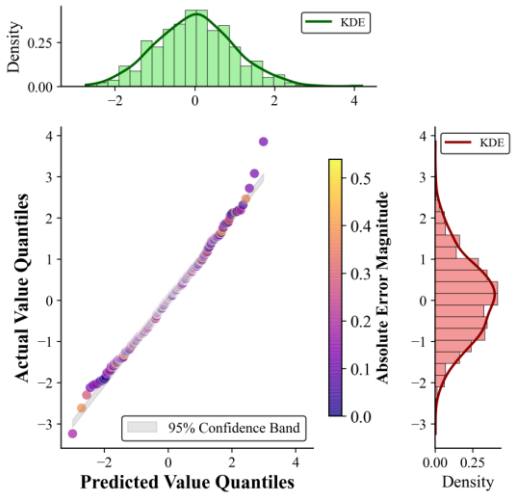

Figure 3: Comparison of prediction error distribution with MAE under different conditions

In Figure 4, the left panel shows the correlation between time intervals T1 and T12, with values ranging from-0.2 to 1.0, and the correlation between T1 and T2 is up to 1.00. The figure on the right shows the correlation between meteorological characteristics. Temperature has a high correlation with humidity, wind speed, cloud cover and other characteristics, with the highest being 1.00 and the lowest being-0.2.

Through correlation analysis, the important correlation between time interval and meteorological characteristics can be identified, which provides a basis for further research.

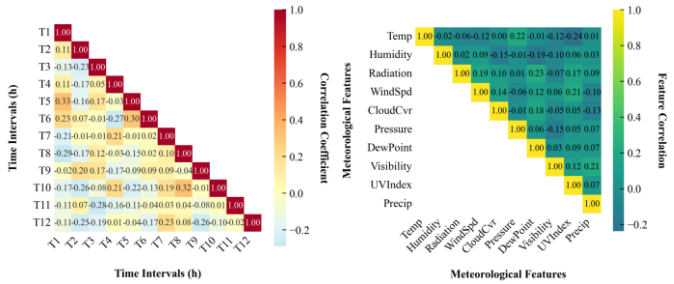

Figure 4: Correlation analysis between time interval and meteorological characteristics

In Figure 5, the left graph shows the relationship between prediction power and residual error, the residual error ranges from-100W to 100W, and the color indicates the error size. The graph on the right shows the residual error distribution for different time periods of the day, and the moving average shows the error trend. The experimental results show that the predicted power is in the range of 0 to 1000W, the residual error is small, and the error density distribution shows that most of the errors are concentrated near zero.

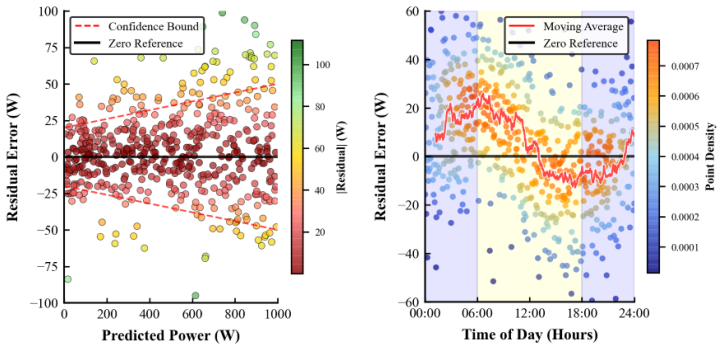

Figure 5: Predicted power error analysis

In Figure 6, the left diagram shows the forecast error distribution under different weather conditions, with errors less than 3% in sunny and partially cloudy conditions, and more than 3% in rainy and snowy days. The upper right figure show the BiLSTM-Informer model has the smallest prediction error.

The lower right graph shows the forecast error for different time periods, with larger errors in rainy days and cloudy conditions.

Experimental values show that the BiLSTM-Informer model shows good prediction performance under different weather and time periods.

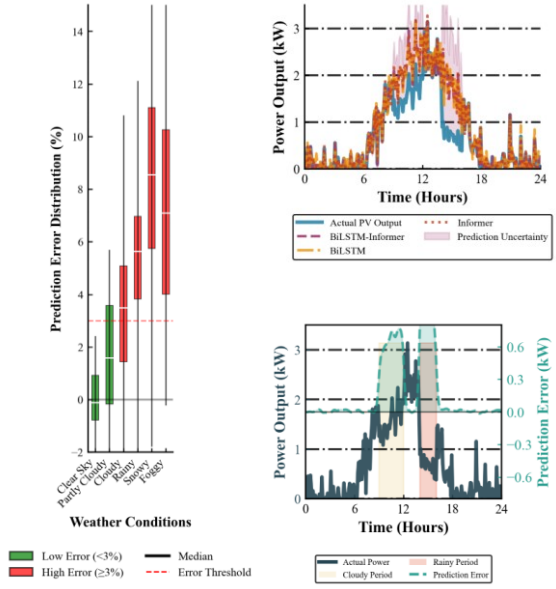

Figure 6: Photovoltaic power generation prediction error analysis

Figure 7's upper panel displays prediction error distributions across various models, while the lower panel presents a box-plot representation of absolute prediction errors. Experimental measurements indicate the BiLSTM architecture achieves minimal error magnitude, the average error is close to 0 kW, and the error distribution is concentrated; BiLSTM-Informer is the second, CNN-LSTM and Informer have larger errors and wide error range; Transformer has the largest error, with an average error of more than 3 kW. The results show that the BiLSTM model performs best in PV power prediction.

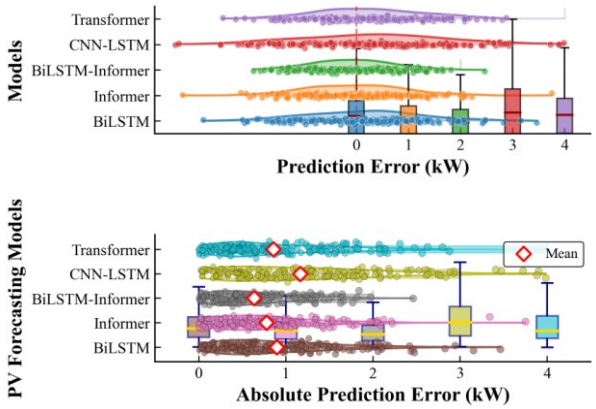

Figure 7: Error comparison of different photovoltaic prediction models

In Figure 8, the left panel shows solar radiation versus normalized PV output with fitted curves R^2 = 0.946 and RMSE = 0.043, indicating a strong linear relationship. The figure on the right shows the relationship between normalized time characteristics and photovoltaic output.

The fitted curve R^2 = 0.848 and RMSE = 0.111 shows a nonlinear relationship. The experimental results show that solar radiation has a significant influence on photovoltaic output, and the influence of time characteristics is complicated.

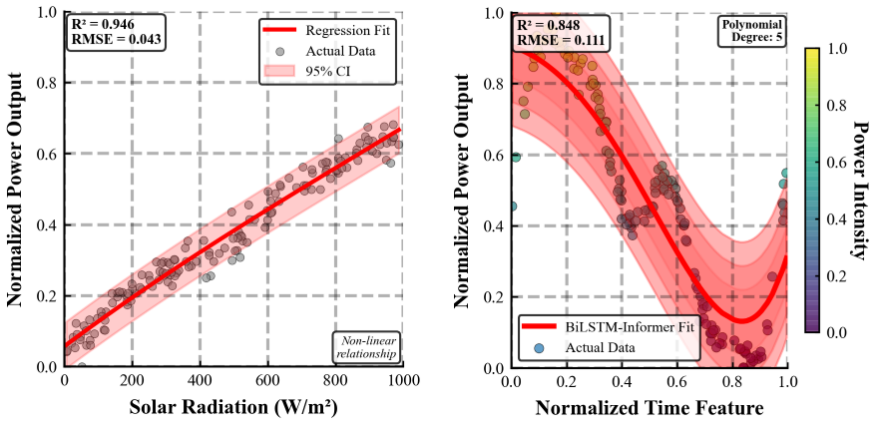

Figure 8: Relationship between photovoltaic output and solar radiation and time characteristics

4. Conclusions

Photovoltaic power generation forecasting needs to take into account long-term dependence capture and real-time requirements. The BiLSTM-Informer hybrid model achieves improvements in prediction accuracy and robustness through architecture fusion and training strategy optimization. The proposed architecture integrates BiLSTM's capacity for local temporal feature extraction with Informer's strength in capturing global dependencies, supplemented by a dynamic training mechanism, to effectively adapt to the fluctuating characteristics of photovoltaic power.

In the model architecture, the BiLSTM layer describes the short-term changes of the power sequence with the help of a bidirectional gating structure, and its output characteristics are linearly transformed and input to the Informer module. The Informer architecture utilizes a ProbSparse attention mechanism alongside encoder distillation, achieving O(L log L) computational complexity for extended sequences while enhancing the representation of global patterns including periodicity and trends.

The model adopts a hierarchical architecture to achieve an effective fusion of local features and global patterns. The composite objective function merges mean square error with L2 regularization, utilizing $\lambda=0.01$ to balance precision against smoothness. During training, stochastic perturbation is introduced through Gaussian noise ($\mu=0$, $\sigma=0.05$), while sequential pattern diversity is enhanced via sliding window operations (window=24, stride=6) to expand the training dataset. A two-stage training strategy is adopted: the BiLSTM components are trained independently first, and then all parameters are jointly optimized. Cosine annealing learning rate scheduling is applied to improve convergence efficiency.

Experiments based on actual operation data of photovoltaic power stations show that the MAE, RMSE and MAPE of the hybrid model reach 0.85, 1.12 and 3.2% respectively, which are better than the baseline model in all aspects. These results verify the effectiveness of the architecture in capturing complex temporal patterns. At the same time, the ProbSparse attention mechanism and hierarchical structure ensure computational efficiency, providing a practical solution for power system scheduling. Subsequent research will focus on the model adaptability under extreme weather conditions and the improvement of online learning strategies.

Funding

Research on Crucial Technologies for the Operati-on of Large-scale Regional Island Power Grids Con-structed upon the "Source-Grid-Load-Storage" Model(2024L02027)

References

[1] Y. Han, X. Hu and K. Li, "Chaotic property based multi-interval informer modeling method for long-term photovoltaic power generation prediction," Applied Soft Computing, vol. 184, no., pp. 113843, 2025. https://doi.org/10.1016/j.asoc.2025.113843

[2] N. K. Rayaguru, N. M. Lindsay, R. G. Crespo and S. P. Raja, "Hybrid bat–grasshopper and bat–modified multiverse optimization for solar photovoltaics maximum power generation," Computers and Electrical Engineering, vol. 106, no., pp. 108596, 2023. https://doi.org/10.1016/j.compeleceng.2023.108596

[3] Z. Wang, Y. Wang, S. Cao, S. Fan, Y. Zhang and Y. Liu, "A robust spatial-temporal prediction model for photovoltaic power generation based on deep learning," Computers and Electrical Engineering, vol. 110, no., pp. 108784, 2023. https://doi.org/10.1016/j.compeleceng.2023.108784

[4] J. Du, J. Zheng, Y. Liang, Q. Liao, B. Wang, X. Sun, H. Zhang, M. Azaza and J. Yan, "A theory-guided deep-learning method for predicting power generation of multi-region photovoltaic plants," Engineering Applications of Artificial Intelligence, vol. 118, no., pp. 105647, 2023. https://doi.org/10.1016/j.engappai.2022.105647

[5] X. Guan, X. Han, J. Wang and T. Wang, "A novel short-term prediction method for distributed photovoltaic power generation considering extreme weather," Engineering Applications of Artificial Intelligence, vol. 162, no., pp. 112540, 2025. https://doi.org/10.1016/j.engappai.2025.112540

[6] J. P. Roth and J. Bajorath, "Unraveling learning characteristics of transformer models for molecular design," Patterns, vol., no., pp. 101392, 2025. https://doi.org/10.1016/j.patter.2025.101392

[7] Y. Guo, X. Lv, G. Yan, S. Chen and S. Di, "TransStyle: Transformer-based StyleGAN for image inversion and editing," Pattern Recognition Letters, vol. 198, no., pp. 1-7, 2025. https://doi.org/10.1016/j.patrec.2025.09.002

[8] W. Wang, Y. Wang, X. Qi and H. Chen, "EIAformer: Empowering transformer with enhanced information acquisition for time series forecasting," Neurocomputing, vol. 658, no., pp. 131700, 2025. https://doi.org/10.1016/j.neucom.2025.131700

[9] D. Zhang, G. Li, B. Ning, Y. Gao, Y. Zhang and D. Yang, "Continuous-Time Transformer with Large Language Model for Temporal Knowledge Graph Forecasting," Knowledge-Based Systems, vol., no., pp. 114695, 2025. https://doi.org/10.1016/j.knosys.2025.114695

[10] R. Wu, Y. Wang, D. Li and J. Liu, "Not all patches are crucial to image recognition: Window patch clustering attention for transformers," Knowledge-Based Systems, vol. 330, no., pp. 114647, 2025. https://doi.org/10.1016/j.knosys.2025.114647

[11] B. Xie, Y. Wang, S. Guo and J. Chen, "Res2former: A multi-scale fusion based transformer feature extraction method," Journal of Visual Communication and Image Representation, vol. 112, no., pp. 104546, 2025. https://doi.org/10.1016/j.jvcir.2025.104546

[12] G. Omondi and T. O. Olwal, "A DNN-based MIMO signal detector using transformer architecture for next-generation wireless networks," Journal of Information and Intelligence, vol., no., pp., 2025. https://doi.org/10.1016/j.jiixd.2025.08.004

[13] J. Miao, M. Zhang and Y. Qiao, "A new multi-object tracking algorithm based on Sparse Detection Transformer," Engineering Applications of Artificial Intelligence, vol. 163, no., pp. 112666, 2026. https://doi.org/10.1016/j.engappai.2025.112666

[14] S. Yang, C. Li, Y. Zhu, H. Shen and L. Fang, "An informer network-based circuit boards fault detection method using infrared temperature series," Microelectronics Reliability, vol. 175, no., pp. 115890, 2025. https://doi.org/10.1016/j.microrel.2025.115890

[15] R. Rastgoo, N. Amjady, S. Islam, I. Kamwa and S. M. Muyeen, "Extreme outage prediction in power systems using a new deep generative Informer model," International Journal of Electrical Power & Energy Systems, vol. 167, no., pp. 110627, 2025. https://doi.org/10.1016/j.ijepes.2025.110627

[16] X. Li, T. Huo, L. Zhu and K. Wang, "Modified BiLSTM network for interval prediction based on Aerospace Load System," Neurocomputing, vol. 651, no., pp. 130887, 2025. https://doi.org/10.1016/j.neucom.2025.130887

[17] Y. Wu, P. Gan, Z. Jing and Q. Yang, "Hybrid CNN-BiLSTM-Attention Conditional GAN for power system forced oscillation localization," International Journal of Electrical Power & Energy Systems, vol. 172, no., pp. 111220, 2025. https://doi.org/10.1016/j.ijepes.2025.111220

[18] W. G. Buratto, R. N. Muniz, A. Nied, C. F. d. O. Barros, E. C. Finardi and G. V. Gonzalez, "Hybrid CF-CNN-BiLSTM hypertuned by Bayesian optimization for thermal power generation and decarbonization forecasting," International Journal of Electrical Power & Energy Systems, vol. 172, no., pp. 111199, 2025. https://doi.org/10.1016/j.ijepes.2025.111199

[19] V. Dubey and R. Katarya, "FloodCNN-BiLSTM: Predicting flood events in urban environments," Engineering Analysis with Boundary Elements, vol. 177, no., pp. 106277, 2025. https://doi.org/10.1016/j.enganabound.2025.106277

[20] A. Abdeltawab, Z. Xi, Z. Longjia and A. M. Galal, "Wavelet-based hybrid CNN-BiLSTM approach in tool wear monitoring," Digital Signal Processing, vol. 168, no., pp. 105529, 2026. https://doi.org/10.1016/j.dsp.2025.105529

[21] K. S. N. Kumari, S. Vijayabaskar, E. Praynlin and K. Aravinda, "Advanced ECG classification with improved optimized feature selection and attention CNN-BILSTM technique," Biomedical Signal Processing and Control, vol. 110, no., pp. 108299, 2025. https://doi.org/10.1016/j.bspc.2025.108299

[22] D. Chi, T. Huang, Z. Jia and S. Zhang, "Research on sentiment analysis of hotel review text based on BERT-TCN-BiLSTM-attention model," Array, vol. 25, no., pp. 100378, 2025. https://doi.org/10.1016/j.array.2025.100378

[23] A. Heidary, M. G. Niasar and M. Popov, "Multi-module series suppressor for the protection of wind farm transformers against resonance overvoltages," International Journal of Electrical Power & Energy Systems, vol. 172, no., pp. 111090, 2025. https://doi.org/10.1016/j.ijepes.2025.111090

[24] S. Lai, J. Wu, W. Wei and C. Ye, "MiKTr: Multiscale knowledge fusion transformer for consumer fraud detection," Information Fusion, vol. 127, no., pp. 103779, 2026. https://doi.org/10.1016/j.inffus.2025.103779

[25] X. Zhuang, J. Ge, X. Mao, D. Zhou, H. Yao, W. Sun, L. Li and J. Xiang, "A novel lightweight model combined with convolutional neural network and transformer for gearbox fault diagnosis using infrared thermal images," Engineering Applications of Artificial Intelligence, vol. 162, no., pp. 112704, 2025. https://doi.org/10.1016/j.engappai.2025.112704

[26] Y. Xu and W. Yang, "Efficient template-separable hierarchical transformer tracking for edge computing," Engineering Applications of Artificial Intelligence, vol. 162, no., pp. 112784, 2025. https://doi.org/10.1016/j.engappai.2025.112784

[27] Q. Li, X. Ren, F. Zhang, L. Gao and B. Hao, "A novel ultra-short-term wind power forecasting method based on TCN and Informer models," Computers and Electrical Engineering, vol. 120, no., pp. 109632, 2024. https://doi.org/10.1016/j.compeleceng.2024.109632

[28] Y. Cui, Z. Li, Y. Wang, D. Dong, C. Gu, X. Lou and P. Zhang, "Informer model with season-aware block for efficient long-term power time series forecasting," Computers and Electrical Engineering, vol. 119, no., pp. 109492, 2024. https://doi.org/10.1016/j.compeleceng.2024.109492

[29] W. Yu, Y. Dai, T. Ren and M. Leng, "Short-time photovoltaic power forecasting based on Informer model integrating Attention Mechanism," Applied Soft Computing, vol. 178, no., pp. 113345, 2025. https://doi.org/10.1016/j.asoc.2025.113345