

Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data

Zehuan Hu ^{a,*}, Yuan Gao ^b, Siyu Ji ^a, Masayuki Mae ^a, Taiji Imaizumi ^a

^a Department of Architecture, Graduate School of Engineering, The University of Tokyo, Tokyo, Japan

^b The Center for Energy Systems Design (CESD), International Institute for Carbon-Neutral Energy Research (WPI-I2CNER); Kyushu University, Japan



ARTICLE INFO

Keywords:

Long-term time series forecasting
Weather forecast data
Long-short term memory
Self-attention
Time series embeddings

ABSTRACT

Accurate predictions of photovoltaic power generation (PV power) are essential for the integration of renewable energy into grids, markets, and building energy management systems. PV power is highly susceptible to weather conditions. Therefore, as weather forecast accuracy improves, it has become increasingly important issue to effectively utilize weather forecast data to enhance prediction accuracy. In this study, an improved model that combines Long Short-Term Memory (LSTM) and self-attention mechanisms is proposed. Proposed model captures the time features through the LSTM network and the correlations among multivariate time series through the self-attention mechanism. Additionally, methods to efficiently integrate historical and forecast data into various time-series forecasting models are also proposed. To verify the effectiveness of the proposed method and the performance of the proposed model, an actual PV power data of a building in Japan is used for various types of experiments. The results demonstrate that the proposed method effectively leverages weather forecast data and enhances the prediction performance of all models, the coefficient of determination (R^2) are improved 15.8% for LSTM model, and 26.4% for proposed model. Whether for short-term or long-term predictions, proposed model consistently provides superior accuracy, practicality, and adaptability across all output sequence lengths. Compared to the basic LSTM model, R^2 on short-term and long-term forecasting increased by 3.9% and 22.5%, respectively.

1. Introduction

1.1. Background

The increasing demand for energy and the depletion of fossil fuels have necessitated the development and utilization of renewable energy sources [1]. Among these sources, solar energy has gained significant traction and has been widely used [2]. However, large-scale photovoltaic grid-connected systems face challenges because of the randomness, volatility, and uncertainty associated with photovoltaic power generation (PV power) [3]. Consequently, accurately forecasting PV power output, mitigating the impact of PV power on grid stability, and promoting the absorption of PV power onto the grid have emerged as important research topics in PV power generation [4].

Advancements in statistics and artificial intelligence technology led to the development of various PV power forecasting methods, which can be categorized into three forecasting time scales: medium and long-term, short-term, and ultra-short-term [5]. Medium and long-term

forecasting is utilized for maintenance planning and annual power generation forecasting [6]. Short-term forecasting is used for dispatching the power grid, improving power supply quality, and incorporating PV power generation into bidding [7]. Ultrashort-term forecasting is employed for dispatching power grids in real-time and ensuring equipment control in PV power generation [8].

PV power forecasting can be classified into three types based on their mechanisms: physical models, statistical methods, and machine learning models [21]. Physical models use a mathematical relationship between the PV power output and solar irradiation; this is calculated using numerical weather prediction or satellite-derived models [22]. Such forecasting considers the tilt and orientation of the PV panels, weather conditions, and location of the PV system. However, the factors affecting physical methods involve complexity and computational burden of the model, especially for high-accuracy models [23].

By applying statistical principles, statistical approaches are capable of extracting correlation and change patterns from historical data.

* Corresponding author.

E-mail address: duankuxiao@gmail.com (Z. Hu).

Table 1
Literature review for PV power prediction.

Ref.	Model	Prediction target	Input parameters
[9]	Physical model	Hourly rooftop PV power	Latitude; longitude, time period; PV panel information; irradiance; temperature; maximum voltage; maximum current; etc.
[10]	Machine learning & optimization algorithm	Hour-ahead PV power	Historical PV power data
[11]	ANN & regression analysis	PV power	Solar irradiation; ambient air and module temperature; wind speed; relative humidity
[12]	LSTM	Day-ahead PV power	Solar irradiance; air temperature; relative humidity; wind speed; cloud; air pressure and weather type
[13]	LSTM	One hour-ahead PV power	Historical PV power data
[14]	LSTM & CNN	One hour-ahead PV power	Historical PV power data and weather data
[15]	LSTM & CNN	Ultra-short-term PV power	Total sky image and weather data
[16]	BiLSTM	Multi-step PV power	Historical PV power data and weather data
[17]	Spatio-temporal graph neural networks	Multi-site PV power prediction	Historical PV power data and weather data
[18]	seq2seq with self-attention	Ultra-short-term PV power	Historical PV power data and weather data
[19]	Dual-Encoder transformer	Short-term PV power	Satellite remote-sensing data and PV station data
[20]	Autoformer	Short-term PV power	Historical PV power data and weather data

Typically, time-series-based forecasting techniques have been developed to capture correlations in PV power curves, including exponential smoothing [24], auto-regressive moving average [25], and auto-regressive integrated moving average [26]. However, time-series data used in statistical methods should be stationary, leading to unsatisfactory PV forecast accuracy on cloudy or rainy days. Statistical methods primarily concentrate on historical data but often overlook weather conditions [27].

In contrast, machine learning methods use the historical data of PV power data and atmospheric variables for directly predicting future PV power time series [28]. These methods can learn patterns and relationships in the data by training large datasets, and they require less input data than physical models [29].

1.2. Deep learning methods for time series forecasting

In recent years, the rapid development of deep learning technology has introduced new opportunities for PV power prediction [30]. Deep learning is an extension of artificial neural networks (ANN), which simulate the learning process of the human brain by constructing multilayered neural networks. Compared to traditional methods, deep learning has a stronger nonlinear modeling and adaptive capabilities, and it can better handle complex time-series data [31]. Recurrent neural network (RNN) is a class of ANN where connections between nodes can create a cycle, and allows output from some nodes to affect subsequent input to the same nodes. Previous research proposed many models based on RNN for time-series forecasting, which includes long short-term memory (LSTM) and gate recurrent unit (GRU) [32]. However, the accuracy of RNN-based models tends to decrease significantly as the length of the prediction sequence increases for medium and long-term prediction problems [33].

A novel model called transformers [34] that uses a self-attention mechanism was proposed for sequence-to-sequence modeling (seq2seq). This model was originally designed for natural language processing applications and generalized to time-series forecasting. Recent years, many models based on the self-attention mechanism were

developed for time-series forecasting [35], including building energy consumption prediction [36], building cooling load prediction [37], solar irradiation prediction [38], electricity price forecasting [39].

1.3. Photovoltaic power forecasting

Several studies demonstrated that LSTM is highly accurate for single-step forecasting of time-series data. Gao et al. [12] proposes day-ahead power output time-series forecasting methods using LSTM networks separately for ideal and non-ideal weather conditions with superior prediction accuracy compared to traditional algorithms. Mohamed et al. [13] advocated for the use of an LSTM-RNN to model temporal variations in PV power, achieving accurate forecasting with an LSTM-based model that outperformed five other evaluated models. This approach resulted in a smaller forecasting error compared to other PV forecasting methods, including multiple linear regression, boosted regression trees, and ANN methods.

Jelena et al. [17] proposed two novel graph neural network models, graph convolutional long-short term memory and dubbed graph convolutional transformer, for multisite PV power forecasting that outperforms the state-of-the-art methods with an average NRMSE error of 8.3% and 8.4% (on the synthetic dataset) and 12.6% and 13.6% (on the real dataset), respectively. Huang et al. [20] proposed a novel transformer model with de-stationary attention and multi-scale framework for short-term PV power forecasting, which incorporates a multi-scale framework and de-stationary attention into an auto-correlation mechanism, and it achieved better performance compared to the six baseline models for predicting short-term PV power output using operational data from a 1058.4 kW PV facility in Central Australia. A summary and review of relevant research on different methods of PV power forecasting is shown in the Table 1.

1.4. Research contribution

Previous research focused on short-term forecasting problems that predicts PV power generation. However, the application of transformer

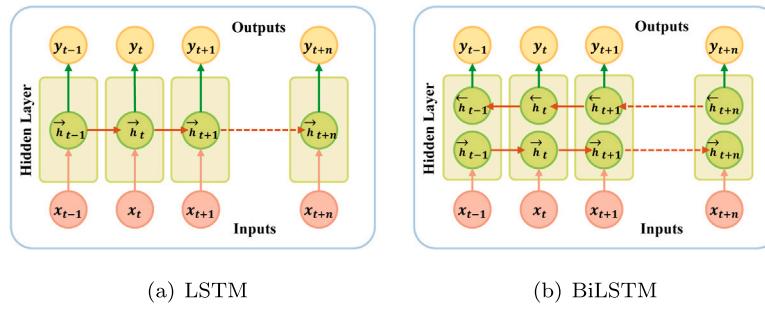


Fig. 1. Topological structures of the LSTM and BiLSTM models.

models to medium and long-term PV power forecasting has been studied insufficiently; the effectiveness of self-attention mechanisms in addressing long-term series forecasting problems requires additional validation. In addition, as the amount of PV power is directly influenced by meteorological conditions, most weather data used in past studies as input sequences are historical data. However, there has been limited research on PV power forecasting performance of RNN-based model and self-attention-based model using weather forecast data as input sequences.

Considering the above limitations in the previous studies, we propose a hybrid model combining LSTM, attention and self-attention, and we study the effect of input and output sequence lengths on different models. The main innovations and contributions of our work are summarized as follows:

- (1) To address the limitations of RNN-based and self-attention models, we propose a novel model based on LSTM model and self-attention mechanism while using a new encoding method to process the input sequence. The proposed model leverages the LSTM network to capture the temporal characteristics of the input data and utilizes the self-attention layers to capture the relationships between feature quantities.
- (2) To enhance the prediction accuracy of time-series forecasting models when the input data is limited, we proposed method that can effectively harness weather forecast data and historical data for various models. This method enables the model to efficiently learn the relationship between weather forecast data and PV power output, allowing the model to achieve high performance even with limited historical data.
- (3) We conduct various experiments using two years and five months of PV power data and measured weather data from an actual building in Japan. The impact of weather forecast data on model prediction performance is analyzed, and the prediction performance of different models is compared under different input sequences and output sequences in our experiments. The experimental results demonstrate the effectiveness of our proposed method and the superiority of the proposed model.

The paper is organized as follows: Methods for all models in this paper are presented in Section 2, including LSTM, BiLSTM, seq2seq transformer, the proposed models, and the proposed method for using weather forecast data. PV power system information, weather data, and experiment settings are presented in Section 3. Section 4 analyzes the effect of weather forecast data on the prediction performance of models, compares the performance of the models, and presents the results and discussion. Finally, Section 5 provides the paper's conclusion.

2. Methodology

2.1. Long-short term memory network (LSTM) and bi-directional LSTM network (BiLSTM)

The LSTM network is an enhanced RNN that comprises a cell, input gate, output gate, and forget gate. The primary function of the

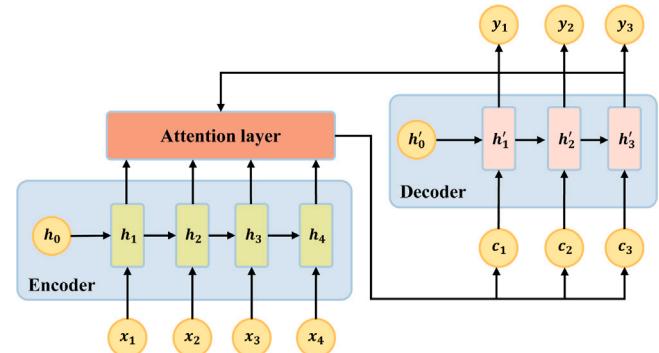


Fig. 2. Topological general structure of sequence-to-sequence model using the attention mechanism.

LSTM layer is acquiring long-term dependencies, and it is employed for predicting time-series data [40]. A fundamental design comprises a collection of LSTM cells and a compact dense output layer (Fig. 1(a)). BiLSTM is a modified version of LSTM and consists of two separate hidden layers. First, the forward hidden sequence is computed, which is followed by the backward hidden sequence, and finally, the two are combined to calculate the output (Fig. 1(b)) [41]. The relevant content and calculation equations of LSTM are provided in the appendix because of space constraints.

2.2. Sequence-to-sequence model with attention (seq2seq)

The sequence-to-sequence model with an attention (seq2seq) model is a network architecture that features an encoder/decoder structure; its input and output are both sequences. The encoder transforms a signal sequence of varying length into a fixed-length context vector, whereas the decoder transforms the fixed-length context vector into a target signal sequence of variable length [42]. Within the conventional seq2seq framework, extracting important information from a long input sequence poses a significant challenge [43]. Yet, the attention mechanism liberates restrictions imposed by the traditional structure, whereby the decoder solely relies on a vector of fixed length. Further, the attention mechanism offers an efficient approach for acquiring a pertinent information. On the one hand, it permits the decoder to retrieve the most relevant data during each decoding step, and on the other hand, it significantly shortens the information flow distance. Consequently, the attention mechanism facilitates the modeling of input and output sequence dependencies without considering their sequence distance, which effectively enhances the distance dependence of the LSTM model [44,45].

The working principle can be observed in Fig. 2. The output of encoder h_t is calculated as

$$h_t = f(x_t, h_{t-1}) \quad (1)$$

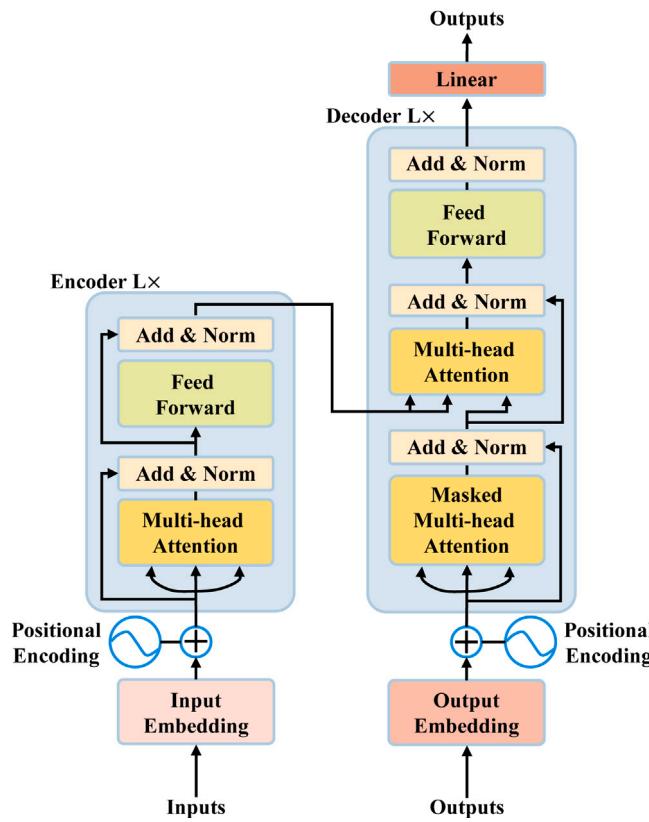


Fig. 3. Topological general structure of the transformer model.

where x_t , h_{t-1} , and f represents the input at time t , hidden state at time $t-1$, and encoder function, respectively. The attention score a_t is calculated as

$$a_t = \text{softmax}(v^T * \tanh(W_1 * h_{t-1} + W_2 * s_t)) \quad (2)$$

where v , W_1 , and W_2 represent the learned parameters, and s_t represents the decoder state at time t . The context vector c_t and s_t is respectively calculated as

$$c_t = \sum_{i=1}^T a_{t,i} * h_i \quad (3)$$

$$s_t = g(y_{t-1}, s_{t-1}, c_t) \quad (4)$$

where T and h_i represent the length of the input sequence and hidden state at time i , and s_{t-1} and g represent the previous decoder state and decoder function. The final output is calculated as

$$y_t = h(s_t) \quad (5)$$

where h represents the output function. In this paper, the functions f and g of the encoder and decoder both use LSTM, and the output function uses a fully connected layer.

2.3. Self-attention transformer model (transformer)

The transformer model (Fig. 3) [34] is a deep learning model architecture that abandons the traditional method that use RNN or convolutional neural network as the encoder and decoder, respectively. Further, it completely uses the attention mechanism to transfer information between different network layers, and it is called self-attention and achieves good results in the natural language processing field. Studies showed that using self-attention instead of CNN or RNN for solving problems related to sequence data not only achieved better

results, but also overcame the problem that RNN cannot be calculated in parallel. In addition, it also speeds up the operation speed [34].

The basic architecture of the self-attention mechanism can be represented by

- (1) Self-attention: A sequence of input embeddings X is transformed into a sequence of output embeddings Z by applying a self-attention mechanism as

$$Q = XW_q, K = XW_k, V = XW_v \quad (6)$$

$$A(Q, K, V) = \text{softmax}(QK^T / \sqrt{d_k})V \quad (7)$$

$$Z = A(Q, K, V)W_o \quad (8)$$

where Q , K , and V represent the query, key, and value matrices respectively; W_q , W_k , W_v , and W_o represent the learnable weights. The attention mechanism is applied to each time step in the input sequence. d_k is the dimension of queries, keys and values, which used to reduce the impact of the input data dimension on the results.

- (2) Positional encoding: The input sequence is augmented with positional encoding vectors for providing information about the position of each time step in the sequence. This is achieved by adding a positional encoding vector to each input embedding.

$$PE_{i,2j} = \sin(i/10000^{2j/d}) \quad (9)$$

$$PE_{i,2j+1} = \cos(i/10000^{2j/d}) \quad (10)$$

where i , j , and d represent the position of the embedding in the sequence, dimension of the embedding, and dimension of the model, respectively.

- (3) Feedforward layer: In the feedforward layer, a feedforward neural network processes the output embeddings Z from the self-attention mechanism to generate the final output of the transformer model.

$$FFN(Z) = \max(0, ZW_1 + b_1)W_2 + b_2 \quad (11)$$

where W_1 , b_1 , W_2 , and b_2 represent learnable weights and biases.

2.4. Proposed model

To address the problem that LSTM has poor prediction accuracy for long time series and the transformer is insensitive to the relationship between features and time, we designed an efficient self-attention-based model for long sequence time series forecasting, named LSTM-transformer. The proposed forecasting model is shown in Fig. 4, and it consists of LSTM encoder, LSTM decoder, attention, data embedding, transformer encoder, and transformer decoder layers based on the self-attention mechanism. The proposed hierarchical architecture allows model to capture the scale-specific inter-and intra-time-series correlations simultaneously by RNN and self-attention module. Specifically, the model first captures the relationship between features and time through the RNN layer in the LSTM Encoder and expands the dimension of the feature, and secondly captures the relationship between each feature through the multi-head self-attention layer in the transformer Encoder. Then the relationship between the output of the LSTM Encoder and the weather forecast information is extracted through the attention mechanism, and the result is passed through the RNN layer again to extract the relationship between the feature and time in LSTM Decoder. Then it goes through the transformer Decoder together with the output of the transformer Encoder, and finally outputs the prediction result through a fully connected layer.

The LSTM models are adept at capturing the time-series pattern through their inherent recurrent structure, with minimal dependence on time stamps [40]. The transformer model relies on a point-wise

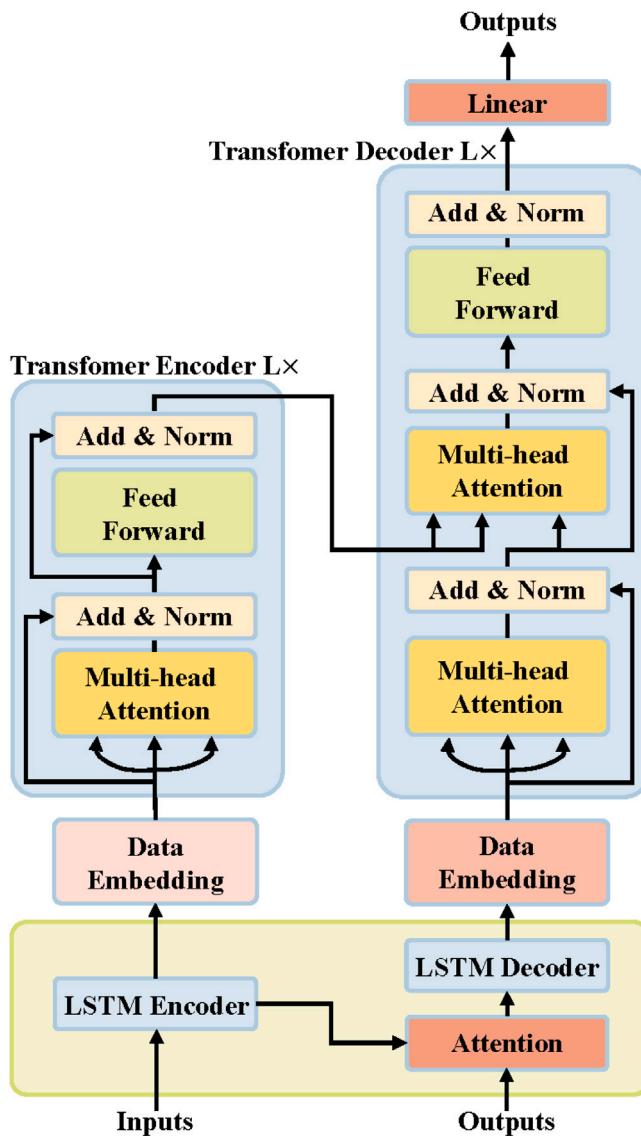


Fig. 4. Topological general structure of the LSTMformer model.

self-attention mechanism, wherein time stamps serve as local positional context [34]. However, in the long sequence time-series forecasting problem, capturing long-range independence necessitates the utilization of global information such as hierarchical and agnostic time stamps. Unfortunately, these are not adequately harnessed in the canonical self-attention mechanism, which leads to query-key mismatches between the encoder and decoder that undermines forecasting performance. To overcome this issue, we propose a uniform input representation, as illustrated in Fig. 5, to provide an intuitive overview of our approach.

A fixed position embedding PE can be calculated by

$$\text{PE}_{pos,2j} = \sin(pos/10000^{2j/d_{model}}) \quad (12)$$

$$\text{PE}_{pos,2j+1} = \cos(pos/10000^{2j/d_{model}}) \quad (13)$$

where $j \in \{1, \dots, d_{model}/2\}$, d_{model} represents the feature dimension after input representation.

We utilize a set of learnable stamp embeddings $\text{SE}_{(pos)}$ with a limited vocabulary size for each global time stamp. This approach allows the self-attention mechanism for accessing global context during similarity computation, while maintaining the computational cost

affordable even for long input sequences. We utilize 1-D convolutional filters (kernel width=3, stride=1) to project the scalar context x_i^t onto a d_{model} -dimensional vector u_i^t for ensuring the dimensionality alignment. Therefore, we obtain the feeding vector as

$$x_i'^t = u_i^t + \text{PE}_{(L_x \times (t-1)+i)} + \sum_p [\text{SE}_{(L_x \times (t-1)+i)}]_p \quad (14)$$

where $i \in \{1, \dots, L_x\}$, and L_x is input sequence length.

2.5. Weather forecast data input method for various models

PV power output is influenced by various weather factors, such as irradiance, temperature, humidity, and more [46]. Brester et al. [47] studied the impact of historical observation data and historical weather forecast data on the accuracy of power generation prediction. The results showed that the model trained using historical numerical weather forecast data has higher accuracy than the model trained using historical observed weather data. Polasek et al. [48] studied the impact of uncertainty in weather forecast data on PV power forecast accuracy. The results showed that allowing the model to evaluate the uncertainty of weather forecasts during the training process can effectively improve forecast accuracy. Hossain et al. [49] proposed a method to predict solar power generation using synthetic weather forecasts generated by K-means classification of historical solar irradiance data. The results showed that the proposed method can effectively improve the prediction accuracy. The above-mentioned research shows that historical data can illustrate the connection between solar power generation and current weather conditions, while forecast data illustrate the future weather changes. Both factors significantly influence the predictive performance of models. However, most of existing studies either exclusively employ historical weather data or rely solely on numerical weather forecast data for model training [50,51]. Few studies explore the integration of both historical and forecast data into conventional prediction models to improve the prediction performance [38].

To address this challenge, we propose a method that can simply input both of historical weather data and weather forecast data into common time series prediction models without any complex preprocessing. In our method, as illustrated in Fig. 6, prediction models can simultaneously learn the impact of historical weather data and weather forecast data on PV power during training. For LSTM and BiLSTM models, the feature dimension of the forecast weather data is reshaped to match the feature dimension of the input data. It is then directly concatenated along the time dimension before being input into the model. In the case of the seq2seq model, the transformer model, and the proposed model, the weather forecast data is processed to have the same dimensions as the output of the encoder, and it is subsequently input into the decoder. In the method we propose, the number of features in historical data and weather forecast data can be either the same or different. For example, it is possible to input all available historical data, while selecting only weather forecast data with minimal errors, such as temperature and humidity.

2.6. Model setup

In the LSTM/BiLSTM/seq2seq models, the number of layers and hidden size significantly affect the fitting of the model. Similarly, the size of the encoder and decoder layers of the self-attention layers in the transformer model has a substantial effect on the fitting of the model. Bayesian Optimization excels compared to other optimizers due to its ability to efficiently find optimal solutions in a limited number of function evaluations, leveraging probabilistic modeling to guide the search in complex and costly objective spaces [52]. Its iterative approach incorporates past evaluations to intelligently select new points for assessment, making it particularly effective in scenarios with expensive or difficult-to-evaluate objective functions [53]. Overall, Bayesian

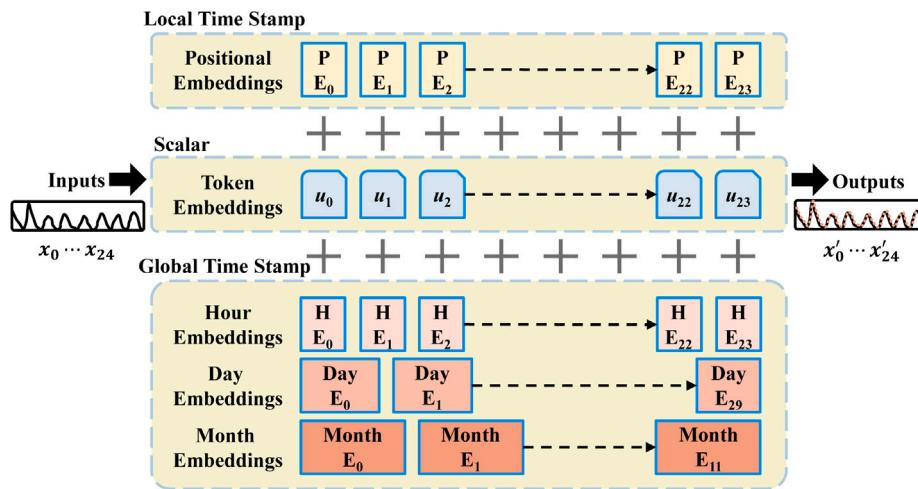


Fig. 5. Data embedding method of the LSTMformer model (the data embedding comprises three separate parts, a scalar projection, the local time stamp, and the global time stamp embeddings).

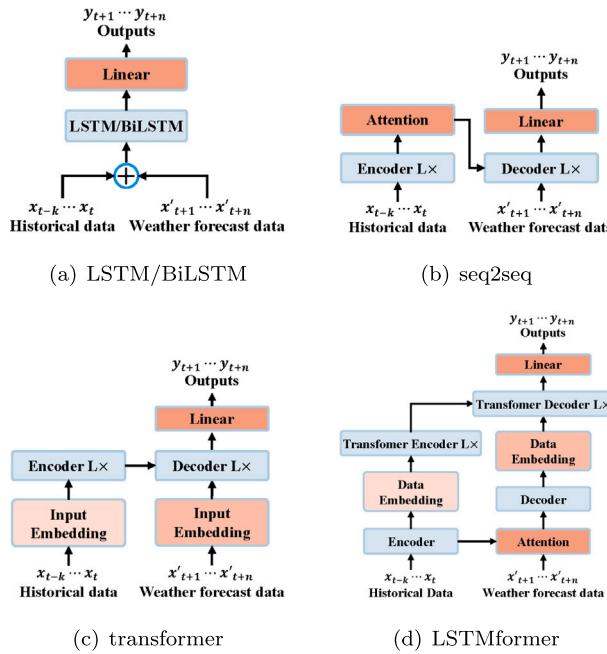


Fig. 6. Input method of historical data and weather forecast data for different models.

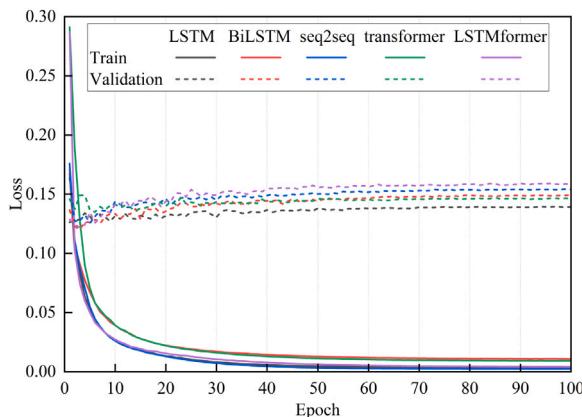


Fig. 7. Loss curves of training and validation loss over epochs.

Table 2

Search space for the hyperparameters of all models and optimizer.

Type	Parameters	Range
Optimizer	Learning rate	[10^{-7} , 10^{-3}]
	Batch size	[8, 64]
LSTM/BiLSTM	Hidden size	[32, 1024]
	Number of layers	[1, 3]
	Fully connected size	[32, 2048]
Seq2seq	Encoder hidden size	[32, 1024]
	Number of encoder layers	[1, 3]
	Decoder hidden size	[32, 1024]
	Number of encoder layers	[1, 3]
	Fully connected size	[32, 2048]
Transformer	Dimensions	[8, 512]
	Number of encoder layers	[1, 12]
	Number of decoder layers	[1, 12]
	Fully connected size	[32, 2048]
LSTMformer	Encoder hidden size	[32, 1024]
	Number of encoder layers	[1, 3]
	Decoder hidden size	[32, 1024]
	Number of encoder layers	[1, 3]
	Dimensions	[8, 512]
	Number of self-attention encoder layers	[1, 12]
	Number of self-attention decoder layers	[1, 12]
Dropout	Fully connected size	[32, 2048]
		[10^{-5} , 0.5]

Optimization is well-suited for tasks like hyperparameters tuning in machine learning, providing a robust and resource-efficient optimization framework [54]. Therefore, in this study, the Bayesian optimizer was used to optimize the hyperparameters, which includes the number of layers, neurons in each layer, activation function, regularization values, batch size, etc., from the search space presented in Table 2 to construct each model. An excessively large RNN network consumes a significant amount of computational resources. Considering the calculation time, we set the maximum number of RNN layers to 3 and the upper limit of size to 1024. Due to the limited amount of data, an overly complex network will not significantly help extract more features, so we set the maximum number of self-attention layers to 12, the maximum size to 512, and the maximum size of the fully connected layer to 2048. The remaining parameters (learning rate, batch size and dropout) are determined through empirical tuning. For each model, the experiments using Bayesian optimization were conducted three times, with the number of trials set to 100. The hyperparameters of each model were determined by selecting the optimal solution among the results of the three experiments. The mean squared error (MSE) loss function is used as the loss function to evaluate the model prediction accuracy.



Fig. 8. PV panel and solar radiation pyranometer in this work.

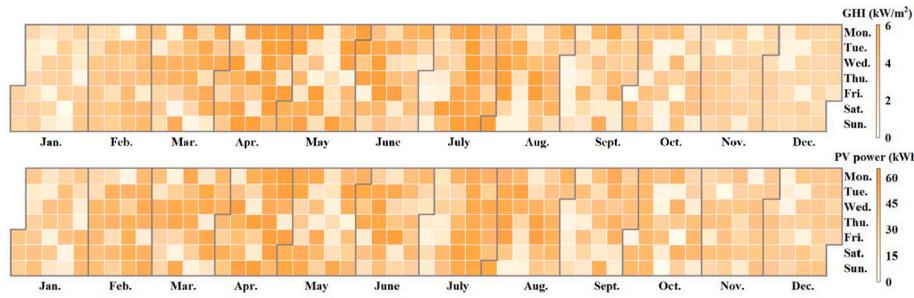


Fig. 9. Average daily global horizontal irradiation (GHI) and PV power for 2021.

The dropout is utilized to prevent overfitting, whereas the learning rate affects the training speed, convergence speed, and overfitting level of the model. In addition, the batch size determines whether the model is prone to local optimization.

An early-stopping scheme is used to prevent over-fitting. This scheme terminates the training process when the accuracy of the validation dataset no longer improves (or begins to decline) after a certain number of iterations. This scheme can expedite the training process and enhance the efficiency of parameter adjustments. The term “patience” in the early stopping mechanism refers to the maximum number of epochs allowed for the performance of the model to not improve; the training is stopped if the model does not improve within that period. In this study, the patience was set to 10, and because of the early-stopping approach, the minimum epochs was set to 100, the loss curve during the training process is shown in Fig. 7. The loss curve indicates that, due to the model’s limited number of trainable parameters, it achieves its best results on the validation dataset within the first ten epochs. However, beyond ten epochs, as the number of epochs increases, the model gradually starts to overfit the training dataset, resulting in an increase in loss on the validation dataset, which eventually stabilizes. In this study, we selected the model that demonstrated the best performance on the validation dataset for further experimentation.

All models were encoded using PyTorch and trained on a computer with a 12th Intel(R) Core(TM) i7-12700F CPU 2.10 GHz and 64 GB of working memory (RAM). The models were solved and calculated using a GPU (NVIDIA GeForce RTX 3080 10 GB).

3. Case study

3.1. Introduction of the dataset

Global horizontal irradiation (GHI) data and PV power data are obtained from the actual measurement of houses in Urayasu City, Chiba

Table 3
PV module specification.

Item	Note
PV module technology	Poly-Si
Power conditioner	11.6 kW
Rated load efficiency	95%
PV capacity	10.5 kW
Azimuth angle	125°
Tilt angle	26°

Prefecture, Japan. The appearance of houses and solar photovoltaic panels are shown in Fig. 8. PV array specifications are reported in Table 3. Fig. 9 shows the measured data of daily GHI and PV power in 2021; this figure shows that PV power has almost the same pattern with GHI during a year.

In addition to GHI and PV power, this paper uses outdoor temperature, outdoor relative humidity (RH), local pressure (LP), sea level pressure (SLP), wind speed (WS), dew point (DP), precipitation, and sunshine duration (SD). Outdoor temperature and RH data are measured by sensors actually installed on the building, and the rest of the data comes from the Japan Meteorological Agency (JMA). The dataset used in the study spans two years and five months from November 2020 to March 2023 and the data interval is one hour, as shown in Table 4. This study is designed to explore the significance of weather forecast data on the accuracy of model predictions and to develop effective methods for integrating weather forecast data into different models. Therefore, we use historical data which is obtained from JMA as weather forecast data and conduct experiments under the assumption that the weather forecast input to the model is correct.

Part of outdoor temperature data, GHI and PV power data are illustrated in Fig. 10. The results indicate the absence of noticeable noise in the data. Furthermore, we consider that the noise within the measured GHI data may encompass information linked to PV power.

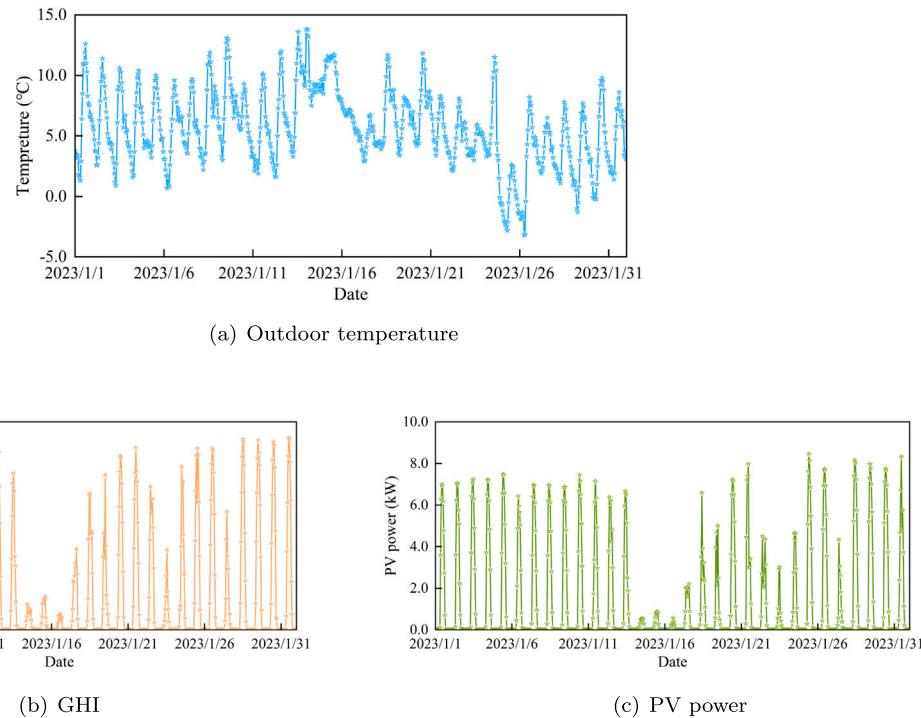


Fig. 10. Outdoor temperature, GHI and PV power data for January 1-31, 2023.

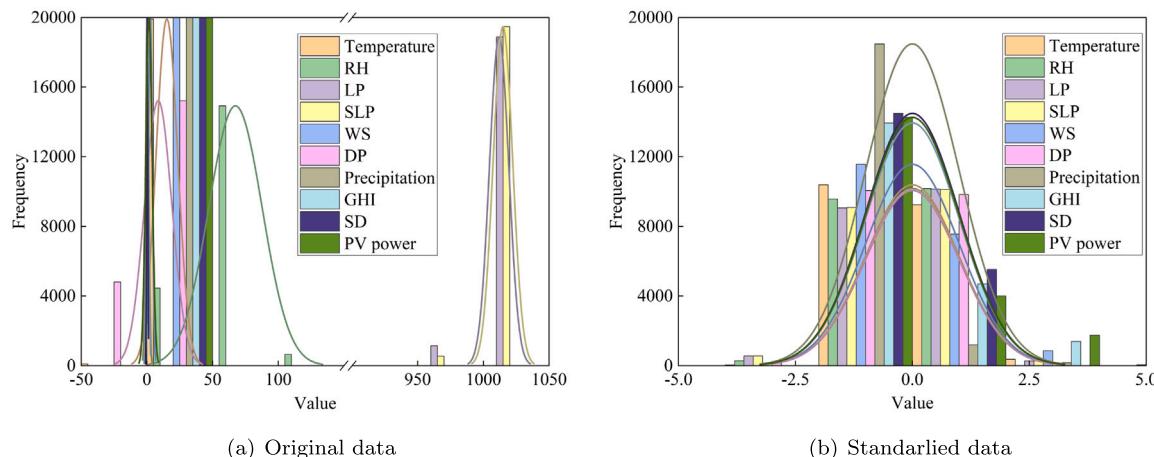


Fig. 11. Data standardization.

Removing this noise could result in information loss. Therefore, we refrain from conducting noise reduction on the data, opting instead to solely remove outliers and interpolate the dataset.

3.2. Data standardization

Data standardization transforms data entries that are originally in different ranges into a consistent range, with a goal reducing regression errors while preserving correlation among the dataset. We utilized Z-score standardization, which re-scales data to have a mean of zero and a standard deviation of one. The formula for Z-score standardization

can be expressed as

$$x' = \frac{x - \mu}{\delta} \quad (15)$$

where x' and x represent the standardized data and original data, respectively; and μ and δ represent the mean and the standard deviation of the original data, respectively. All feature data should be standardized before transmitting to models. The original and standardized data distributions of each feature quantity are presented in Fig. 11.

3.3. Data splitting

Cross-validation is a powerful technique that enhances the reliability of predictive models. The original dataset is partitioned into

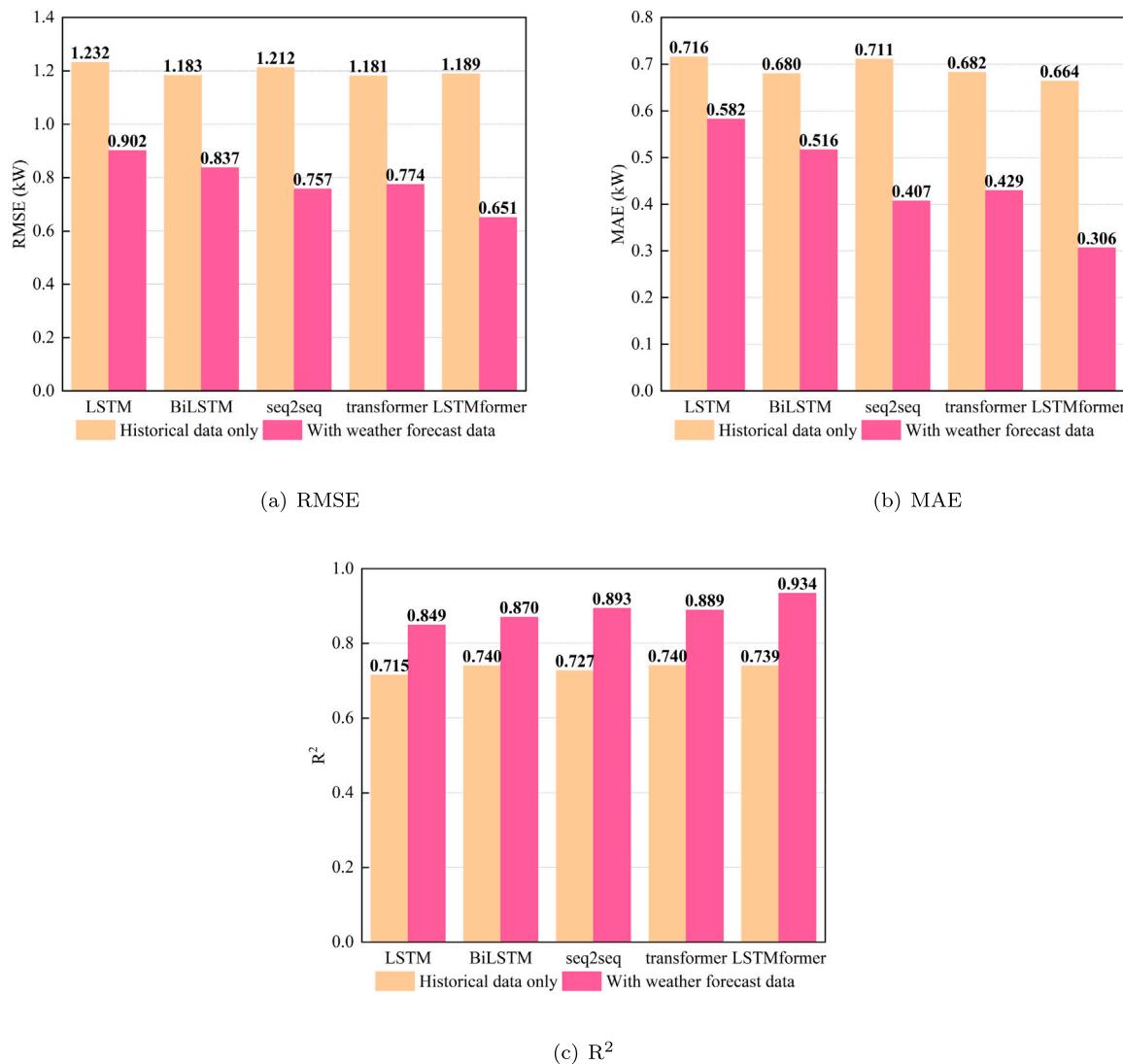


Fig. 12. Variation of the three evaluation indices of the five models using only past data and historical data & weather forecast data as input sequence.

Table 4

Description of the dataset (data interval: one hour; data period: 2020.11.09-2023.03.16; data type: time-series numerical; *: forecast weather data).

Features	Data source	Unit
PV power	PV panel at actual building	kW
Global horizontal irradiation	Pyranometer sensor at actual building	kW/m ²
Outdoor temperature	Temperature sensor at actual building	°C
Outdoor relative humidity	Humidity sensor at actual building	%
Local pressures		hPa
Sea level pressure		hPa
Wind speed		m/s
Dew point		°C
Precipitation	Japan Meteorological Agency	mm
Sunshine duration		h
Outdoor temperature*		°C
Outdoor relative humidity*		%

three subsets, i.e., the training, cross-validation, and test datasets. The training dataset, which comprises a significant portion of the data, is utilized to train the model, whereas the test dataset is employed for evaluating the model and prevent overfitting. In the source domain, we designated 20 months of historical data as the training dataset, 3 months of historical data as the cross-validation dataset, and 6 months of data as the testing dataset.

3.4. Evaluation metrics

We incorporated three widely used criteria to assess the predictive performance of the model from multiple perspectives: root mean squared error (RMSE), mean absolute error (MAE), and coefficient of determination (R^2). These evaluation metrics enable us to gauge the predictive ability of the model from various angles. The formulas to

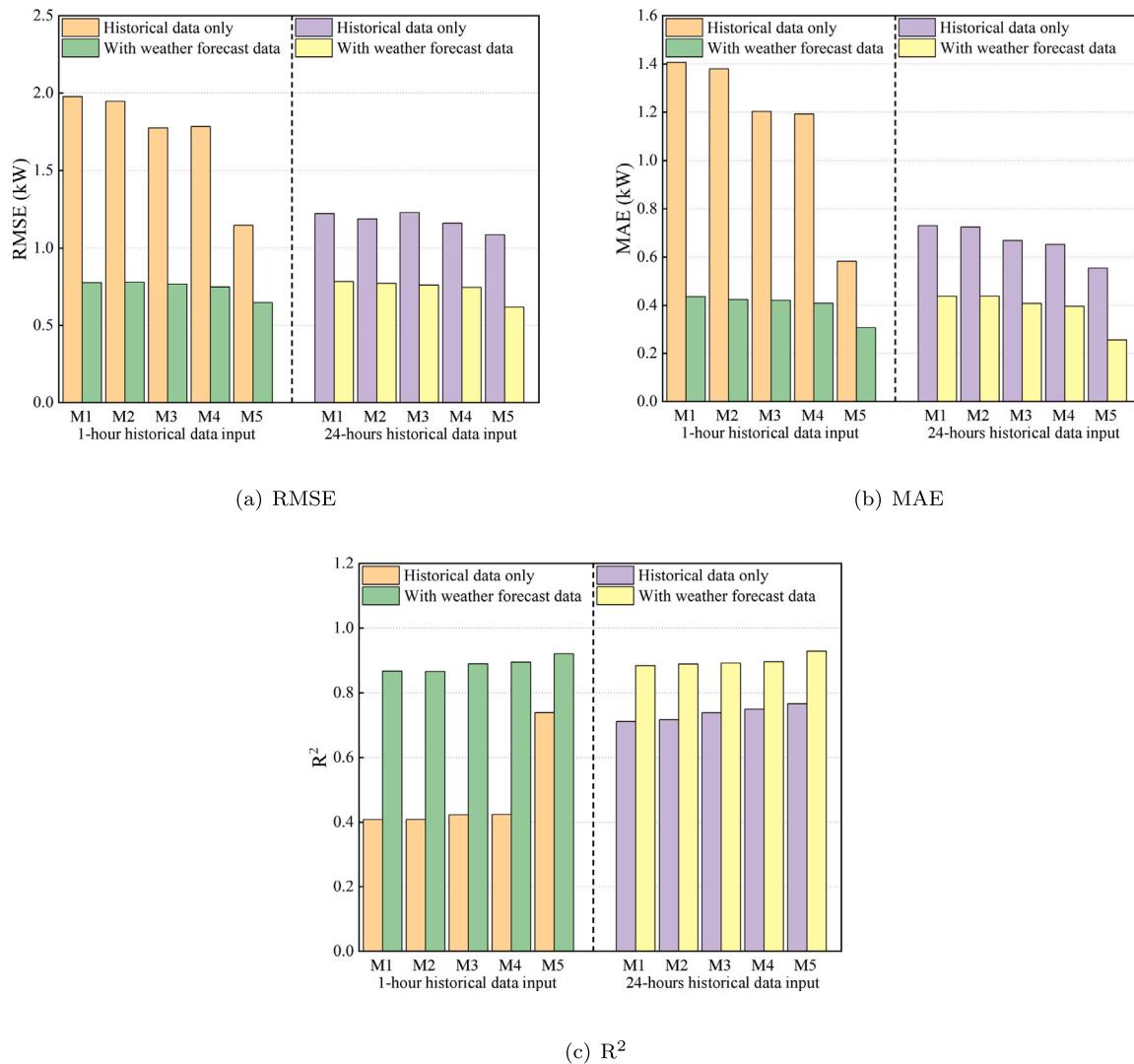


Fig. 13. Comparison of past data and historical data & weather forecast data as input sequences with different input sequence length (M1, M2, M3, M4, and M5 are the LSTM, BiLSTM, seq2seq, transformer, and LSTMformer models, respectively).

calculate RMSE, MAE, and R^2 are

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (16)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (17)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

where y_i , \hat{y}_i , and \bar{y} represent the measured value, forecasted value, and mean of measured value, respectively. As all input data is standardized beforehand, the output of the model represents the predicted value after standardization. We de-standardize the predicted value output by the model and compute the evaluation metrics based on the measured values to more intuitively compare the performance of the model. The formula for inverse standardization can be expressed as

$$x = x' \times \delta + \mu \quad (19)$$

4. Results and discussions

4.1. Effect of weather forecast data on forecast results

The impact of weather forecast data on the prediction accuracy of the model was determined through experiments. In our study, the temperature, humidity, and wind speed, which are obtained easily in the weather forecast, are selected as weather forecast data. The input and output sequence lengths in this experiment are selected to be 72 and 24, respectively. The prediction results (Fig. 12) indicate that, for all five models, inputting known future meteorological parameters can effectively improve the prediction accuracy of the model. Thus, in practical applications, the use of reliable weather forecast data can significantly enhance the prediction accuracy of the model. For the experiments described in the remainder of the article, a combination of past meteorological parameters and weather forecast data was used as input for the model.

We also conducted a comparison between the effect of weather forecast data as input sequence and difference lengths of input sequences on the prediction performance of the model, when the output sequence length was fixed to 72. The results are depicted in Fig. 13. The results showed that, with or without weather forecast data, the prediction performance of the model worsens with a decrease in the length of the

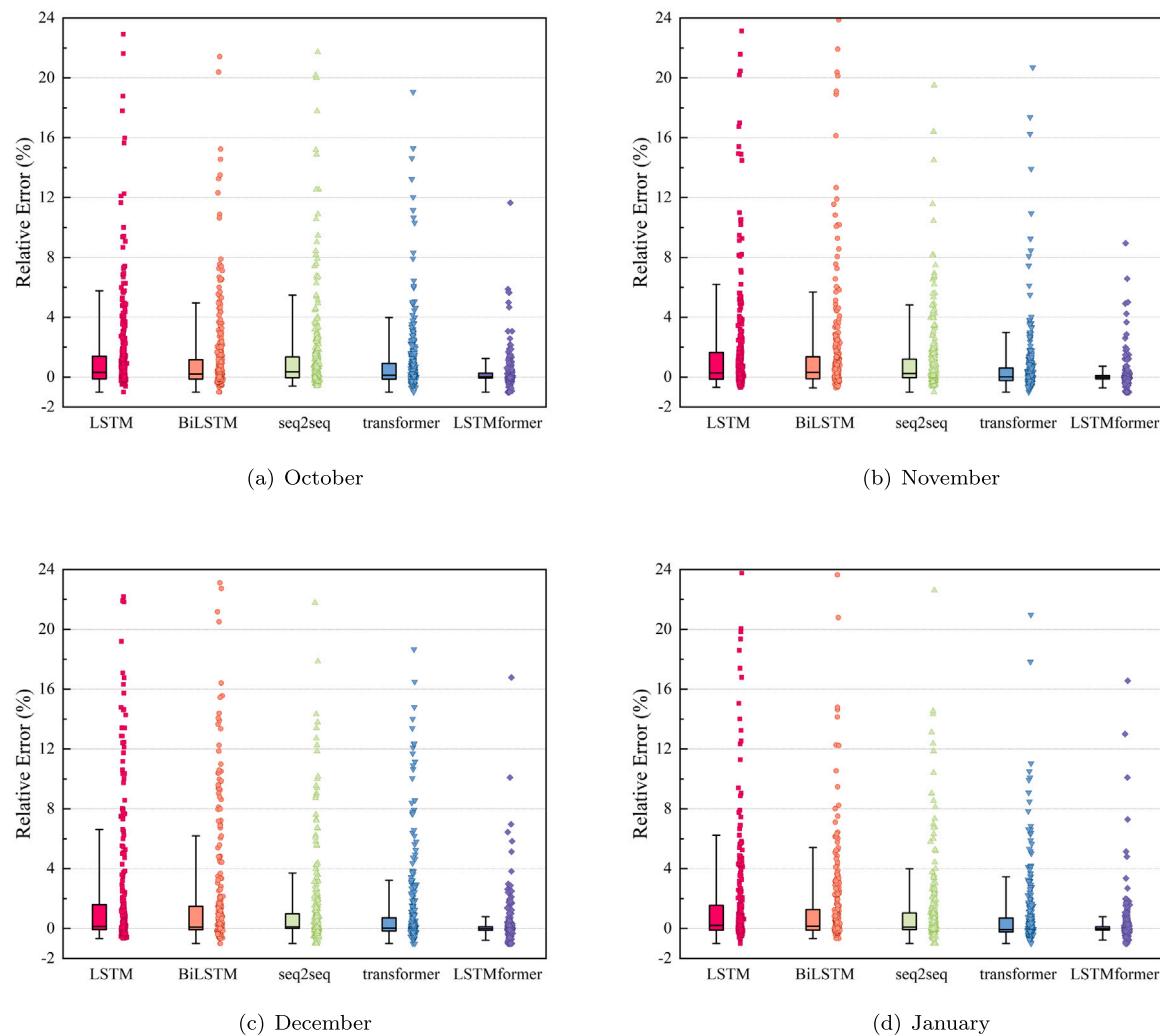


Fig. 14. Relative errors values of the prediction error distribution of the five models on different months.

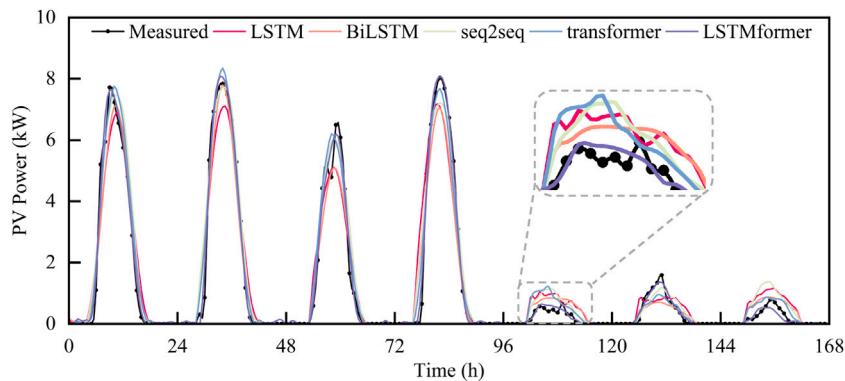


Fig. 15. Comparison between the measured and predicted values of five models for the multi-step ahead PV power prediction.

input sequence. However, with the inclusion of weather forecast data in the input sequence, the performance of the model remains satisfactory even with a very short input sequence length. Thus, the addition of weather forecast data can significantly enhance the accuracy of the model when the available historical data for PV power forecasting is limited.

4.2. Prediction results of five models

Table 5 lists the evaluation metrics of the five models for solar power generation forecasting when the input and predicted sequence lengths are 72 and 24, respectively. The model based on the self-attention mechanism has a better predictive effect than the ordinary

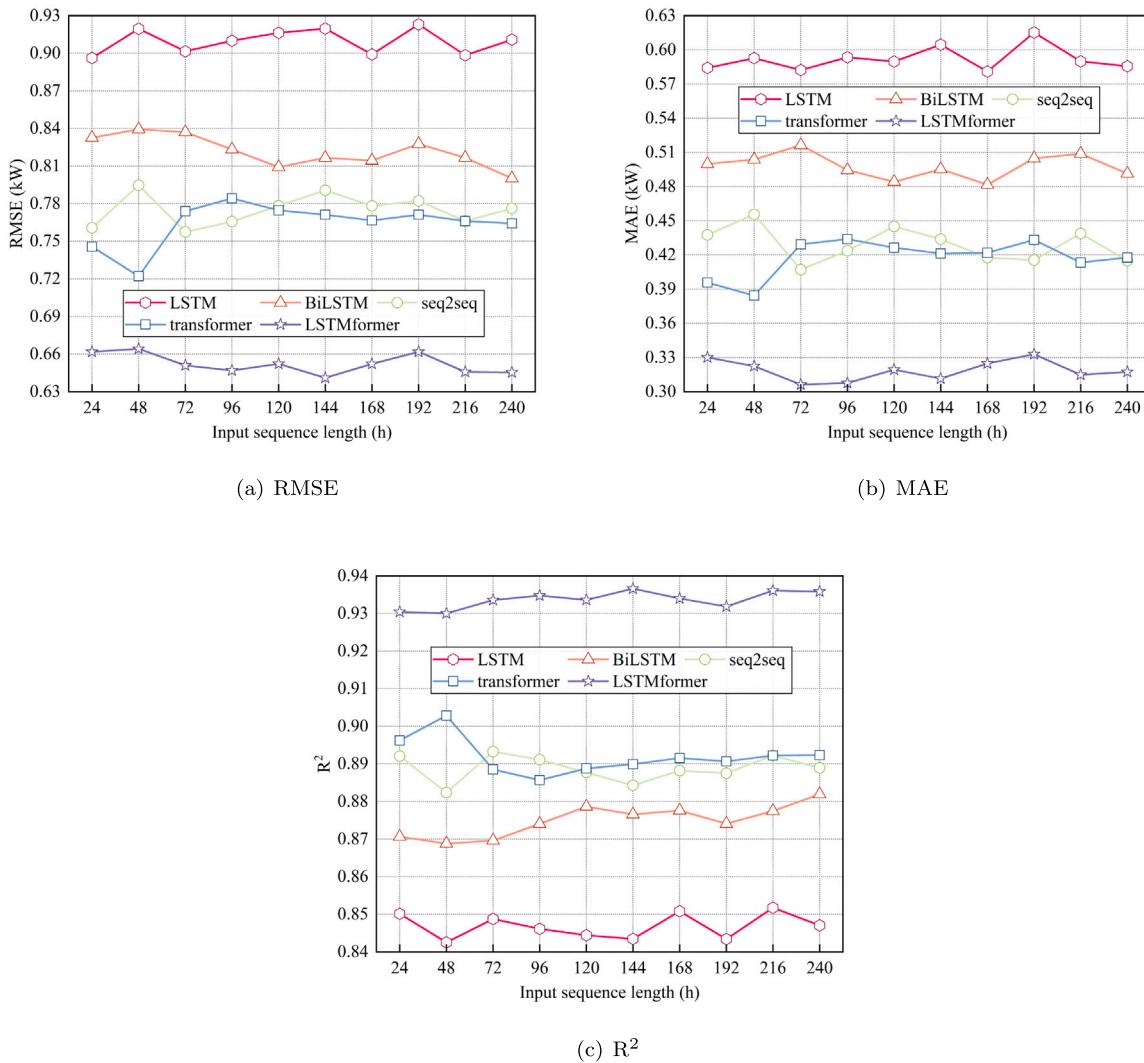


Fig. 16. Variation of the three evaluation indices of the five models at different input sequence lengths.

Table 5
Prediction results of five models.

Model	Metrics		
	RMSE (kW)	MAE (kW)	R ²
LSTM	0.902	0.582	0.849
BiLSTM	0.837	0.516	0.870
seq2seq	0.757	0.407	0.893
transformer	0.774	0.429	0.889
LSTMformer	0.651	0.306	0.934

RNN model. The simple LSTM model performed the worst in prediction, with RMSE, MAE, and R² of 0.902 kW, 0.582 kW, and 0.849, respectively. The BiLSTM model is better than the LSTM model, and its RMSE, MAE, and R² are 0.837 kW, 0.516 kW, and 0.870, respectively. By introducing the attention mechanism in the RNN-based model can effectively improve prediction accuracy. The seq2seq model with attention has reached an accuracy similar to the transformer model and RMSE, MAE, and R² are 0.757 kW, 0.407 kW, and 0.893, respectively. The proposed LSTMformer model has the best prediction effect, and RMSE, MAE, and R² are 0.651 kW, 0.306 kW, and 0.934, respectively, compared with the ordinary LSTM model; the prediction effect is improved by 51.9%, 27.8%, 47.4%, and 10.0%.

In Fig. 14, the boxplot shows the PV power forecasting performance and distribution of the relative error in different months, whereas the point marks indicate the outliers of the extreme relative prediction error. The RNN-based models have greater variations than those of the self-attention-based models; the proposed LSTMformer model has the lowest errors and the smallest variation around the mean over four months of the test dataset.

Fig. 15 shows the comparison between the predicted and actual results of the different models for solar power production over a week in the test set that included sunny and cloudy days. The figure shows that both the RNN-based and self-attention-based models can better simulate the changing trends of solar power generation. The proposed LSTMformer model has the highest forecasting accuracy and can predict solar power generation even under highly variable weather conditions.

4.3. Influence of input sequence length

The article compares the performance of model predictions with input sequence lengths ranging from 24 to 240 to investigate the influence of input sequence length on model prediction accuracy, while keeping the output sequence length fixed at 24. According to Fig. 16, the prediction performance of the five models is relatively stable across

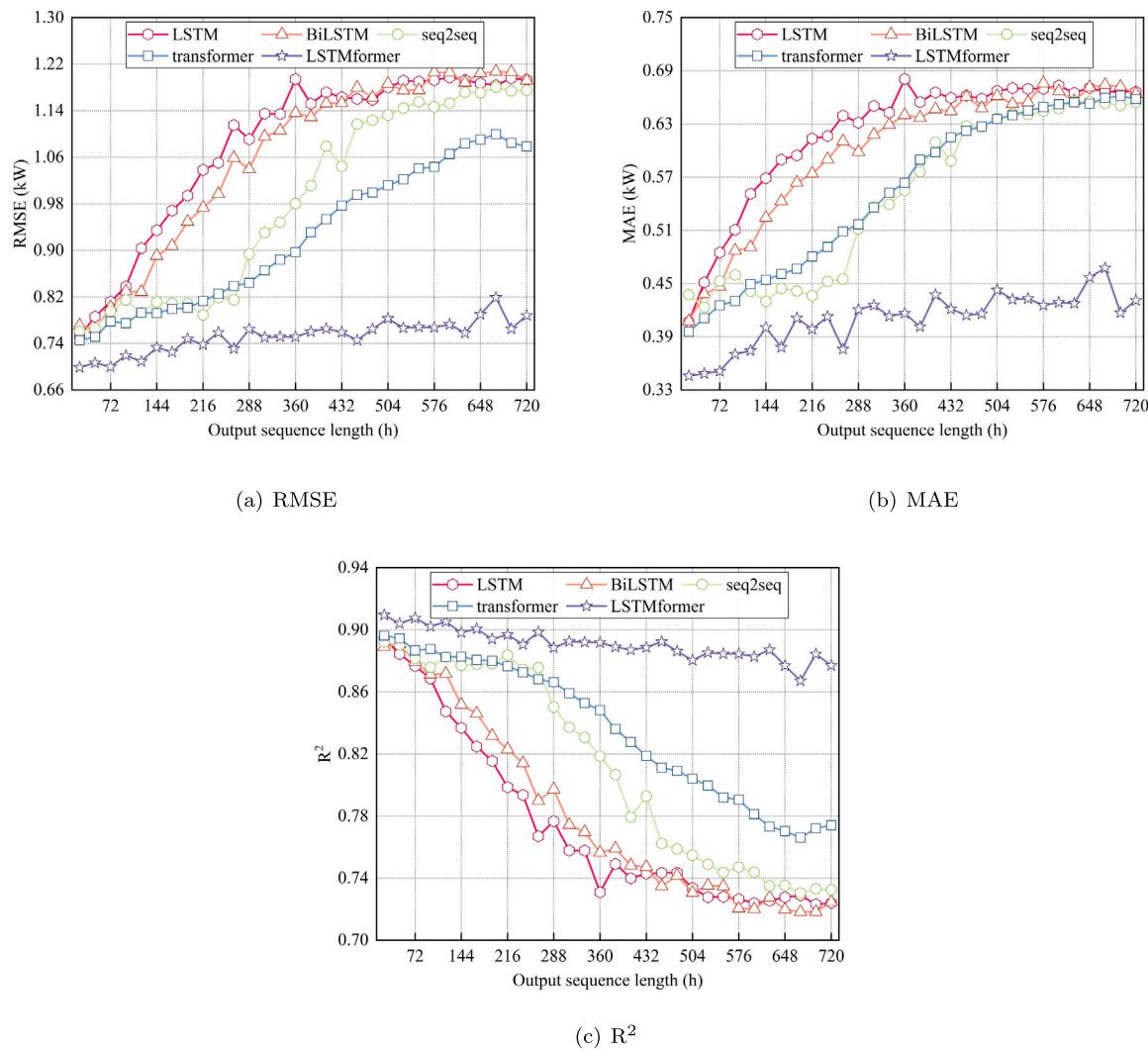


Fig. 17. Variation of the three evaluation indexes of the five models at different output sequence lengths.

varying input sequence lengths and all models demonstrate good accuracy when at least one full day of data is input.

4.4. Influence of prediction sequence length

Fig. 17 illustrates changes in the prediction performance of the five models with an increasing length of the prediction sequence, which is ranging from 24 to 720. The R^2 of the five models decreased with an increase in the length of the output time steps, while RMSE and MAE increased. This indicates that the prediction performance of the models gradually declined with an increasing number of prediction time steps because the correlation between historical and target values weakened over time. Owing to their limited ability to effectively capture distant information, the LSTM and BiLSTM models display the fastest decline in predictive performance. The predictive performance of the seq2seq model declines relatively slowly because the introduced attention mechanism helps capture mid-range information attention, albeit with only moderate performance in capturing longer-distance information. In comparison, the transformer model, which uses a self-attention mechanism, shows slower performance degradation than that of the RNN-based model, indicating that the self-attention mechanism significantly enhances its ability to capture long-distance information. The proposed LSTMformer model combines both LSTM and self-attention mechanisms, displaying the best prediction stability and slowest performance decline among the five models, while being able to effectively

capture both short and long-distance information. Compared with the LSTM model, RMSE, MAE, and R^2 are improved from 1.082, 0.621, and 0.774 to 0.710, 0.367, and 0.901 on the average of all output sequence lengths.

Analysis of the prediction results from the five models reveals that, when compared to the basic LSTM, the BiLSTM can capture input sequence features both in forward and backward time order, thereby enhancing the accuracy of long-term series forecasting to some extent. Incorporating an attention mechanism into the LSTM model results in a more effective capture of connections among input sequence features, leading to further improvements in prediction accuracy. Compared to the standard attention mechanism, the multi-head self-attention mechanism in the transformer excels at capturing and extracting various features of input sequence relationships repeatedly, leading to more effective results than traditional attention mechanisms. However, the multi-head self-attention in the transformer captures the connections among all features simultaneously, rather than extracting features in a time series fashion as in the LSTM network. This limitation means it loses some ability to capture time-related features within the sequence. The proposed hybrid model possesses the capability to simultaneously extract various features of the input sequence through multi-head self-attention and LSTM for capturing temporal features. This combination yields the best results for both short-term and long-term forecasting.

5. Conclusion

We proposed a model that combines LSTM, an attention mechanism, and a self-attention mechanism based on their respective characteristics to capture the time features among variables and the correlations among multivariate time series simultaneously. We also proposed a method to use historical and weather forecast data as input into different models. We analyzed the impact of weather forecast data on the performance of models, and the effects of input and output sequence lengths on the performance of the model. Our findings are listed below.

- (1) The proposed method can effectively utilize weather forecast data to enhance prediction accuracy for all models. The results indicated that compared to using only historical data as input, the R^2 of the LSTM, BiLSTM, seq2seq, transformer and the proposed LSTMformer model is improved by 15.8%, 17.6%, 22.8%, 20.1% and 26.4%, respectively.
- (2) When using only historical data, the model's prediction accuracy is highly sensitive to the length of the input sequence. The prediction R^2 when inputting 1-hour historical data and inputting 24-hour historical data can differ by as much as 76.8%. However, when using weather forecast data, all models maintain high prediction accuracy even with minimal historical data provided. The results demonstrate that, when employing the proposed method, the R^2 of the LSTMformer model can reach 0.921 with just 1-hour of historical data as input.
- (3) The RNN-based models perform poorly in long-term time series prediction, and in contrast, the self-attention based models outperform RNN-based models in long-term time series prediction. The proposed model demonstrates exceptional accuracy, practicality, and adaptability in all scenarios. The results indicate that the R^2 of the proposed model only decreased by 3.6% when the R^2 of the LSTM model decreased by 19.0%.

In this paper, we used actual values obtained from JMA as weather forecast data, so we did not examine the effect of biased future weather forecast data on the performance of the model. In future studies, we aim to investigate how the accuracy of future weather forecasts affects the performance of the model. In addition, we utilized all meteorological data available from JMA as features for the input sequence. In future, we plan to explore the effect of using only a limited number of features as input on the performance of the model.

CRediT authorship contribution statement

Zehuan Hu: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Yuan Gao:** Validation, Methodology, Conceptualization. **Siyu Ji:** Visualization, Formal analysis, Data curation. **Masayuki Mae:** Writing – review & editing. **Taiji Imaizumi:** Writing – review & editing, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

This work was supported by project No. 23KJ0766 funded by the Japan Society for The Promotion of Science.

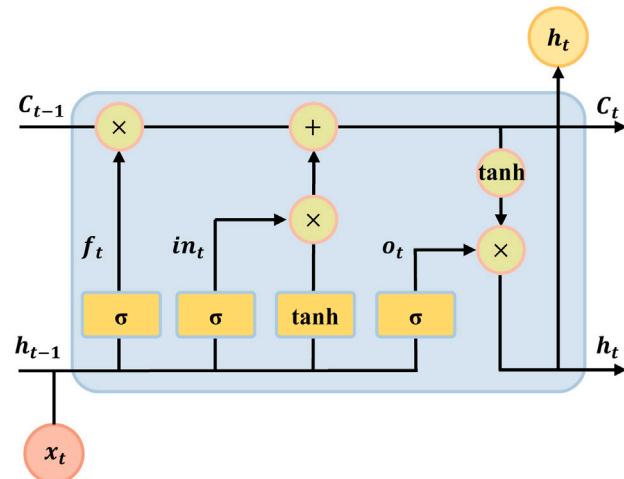


Fig. 18. Unfold structure of the LSTM unit.

Appendix. Long short-term memory network

The essence of the LSTM unit resides in its trio of gating units regulated by the sigmoid function. The unit can calculate and sift through the proportion of past and current input information using these sigmoid functions, and this culminates in the output of the final outcome for the current time step via the output gate (as depicted by h_t in Fig. 18). The cell state c_t in the diagram is impervious for the gating unit and remains relatively constant throughout the calculation process. This conveyor belt-style computing system serves as a vital component in preserving the long-term memory of the LSTM network.

LSTM is calculated using:

$$gate^f = \sigma(b_f + U_f \cdot x_t + W_f \cdot h_{t-1}) \quad (20)$$

$$gate^i = \sigma(b_i + U_i \cdot x_t + W_i \cdot h_{t-1}) \quad (21)$$

$$gate^o = \sigma(b_o + U_o \cdot x_t + W_o \cdot h_{t-1}) \quad (22)$$

$$\tilde{c}_t = \tanh(b_c + U_c \cdot x_t + W_c \cdot h_{t-1}) \quad (23)$$

$$c_t = gate^f \odot c_{t-1} + gate^i \odot \tilde{c}_t \quad (24)$$

$$h^t = gate^o \odot \tanh(c^t) \quad (25)$$

where $gate^f$, $gate^i$, and $gate^o$ represents the calculation results of the forget gate, input gate, and output gate, respectively; W , U , and b refer to the weights and biases in the model; \odot and σ are the Hardmdard product and sigmoid function.

References

- [1] Irena T. Renewable energy statistics 2020. The International Renewable Energy Agency (IRENA) Abu Dhabi; 2020.
- [2] Kabir E, Kumar P, Kumar S, Adelodun AA, Kim K-H. Solar energy: Potential and future prospects. Renew Sustain Energy Rev 2018;82:894–900.
- [3] Maghami MR, Hizam H, Gomes C, Radzi MA, Rezadad MI, Hajighorhani S. Power loss due to soiling on solar panel: A review. Renew Sustain Energy Rev 2016;59:1307–16.
- [4] Saber EM, Lee SE, Manthapuri S, Yi W, Deb C. PV (photovoltaics) performance evaluation and simulation-based energy yield prediction for tropical buildings. Energy 2014;71:588–95.
- [5] Han S, Qiao Y-h, Yan J, Liu Y-q, Li L, Wang Z. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl Energy 2019;239:181–91.

- [6] Moreira M, Balestrassi P, Paiva A, Ribeiro P, Bonatto B. Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting. *Renew Sustain Energy Rev* 2021;135:110450.
- [7] Yin W, Han Y, Zhou H, Ma M, Li L, Zhu H. A novel non-iterative correction method for short-term photovoltaic power forecasting. *Renew Energy* 2020;159:23–32.
- [8] Zheng L, Liu Z, Shen J, Wu C. Very short-term maximum Lyapunov exponent forecasting tool for distributed photovoltaic output. *Appl Energy* 2018;229:1128–39.
- [9] Zhi Y, Sun T, Yang X. A physical model with meteorological forecasting for hourly rooftop photovoltaic power prediction. *J Build Eng* 2023;106997.
- [10] Wang J, Zhou Y, Li Z. Hour-ahead photovoltaic generation forecasting method based on machine learning and multi objective optimization algorithm. *Appl Energy* 2022;312:118725.
- [11] Keddouda A, Ihaddadene R, Boukhari A, Atia A, Arici M, Lebbihiat N, et al. Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Convers Manage* 2023;288:117186.
- [12] Gao M, Li J, Hong F, Long D. Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using LSTM. *Energy* 2019;187:115838.
- [13] Abdel-Nasser M, Mahmoud K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput Appl* 2019;31:2727–40.
- [14] Xiao Z, Huang X, Liu J, Li C, Tai Y. A novel method based on time series ensemble model for hourly photovoltaic power prediction. *Energy* 2023;276:127542.
- [15] Bo Y, Xia Y, Ni Y, Liu K, Wei W. The ultra-short-term photovoltaic power prediction based on multi-exposure high-resolution total sky images using deep learning. *Energy Rep* 2023;9:123–33.
- [16] Lin W, Zhang B, Li H, Lu R. Multi-step prediction of photovoltaic power based on two-stage decomposition and BiLSTM. *Neurocomputing* 2022;504:56–67.
- [17] Simeunović J, Schubnel B, Alet P-J, Carrillo RE. Spatio-temporal graph neural networks for multi-site PV power forecasting. *IEEE Trans Sustain Energy* 2021;13(2):1210–20.
- [18] Ju Y, Li J, Sun G. Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning. *IEEE Access* 2020;8:44821–9.
- [19] Cao H, Yang J, Zhao X, Yao T, Wang J, He H, et al. Dual-encoder transformer for short-term photovoltaic power prediction using satellite remote-sensing data. *Appl Sci* 2023;13(3):1908.
- [20] Huang Y, Wu Y. Short-term photovoltaic power forecasting based on a novel autoformer model. *Symmetry* 2023;15(1):238.
- [21] Das UK, Tey KS, Seyedmahmoudian M, Mekhilef S, Idris MYI, Van Deventer W, et al. Forecasting of photovoltaic power generation and model optimization: A review. *Renew Sustain Energy Rev* 2018;81:912–28.
- [22] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl Energy* 2021;283:116239.
- [23] Ahmed R, Sreeram V, Mishra Y, Arif M. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew Sustain Energy Rev* 2020;124:109792.
- [24] Tratar LF, Strmčník E. The comparison of holt-winters method and multiple regression method: A case study. *Energy* 2016;109:266–76.
- [25] Chu Y, Urquhart B, Gohari SM, Pedro HT, Kleissl J, Coimbra CF. Short-term reforecasting of power output from a 48 MWe solar PV plant. *Sol Energy* 2015;112:68–77.
- [26] Bouzerdoum M, Mellit A, Pavan AM. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Solar Energy* 2013;98:226–35.
- [27] Tang Y, Yang K, Zhang S, Zhang Z. Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy. *Renew Sustain Energy Rev* 2022;162:112473.
- [28] Zhou Y, Liu Y, Wang D, Liu X. Comparison of machine-learning models for predicting short-term building heating load using operational parameters. *Energy Build* 2021;253:111505.
- [29] Dosdoğru AT, İpek A. Hybrid boosting algorithms and artificial neural network for wind speed prediction. *Int J Hydrogen Energy* 2022;47(3):1449–60.
- [30] Wang K, Qi X, Liu H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl Energy* 2019;251:113315.
- [31] Wang F, Zhen Z, Mi Z, Sun H, Su S, Yang G. Solar irradiance feature extraction and support vector machines based weather status pattern recognition model for short-term photovoltaic power forecasting. *Energy Build* 2015;86:427–38.
- [32] Wang Y, Liao W, Chang Y. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies* 2018;11(8):2163.
- [33] Venkatraman A, Hebert M, Bagnell J. Improving multi-step prediction of learned time series models. In: Proceedings of the AAAI conference on artificial intelligence, vol. 29, no. 1. 2015.
- [34] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. *Adv Neural Inf Process Syst* 2017;30.
- [35] Lim B, Arik SÖ, Loeff N, Pfister T. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *Int J Forecast* 2021;37(4):1748–64.
- [36] Gao Y, Ruan Y. Interpretable deep learning model for building energy consumption prediction based on attention mechanism. *Energy Build* 2021;252:111379.
- [37] Li A, Xiao F, Zhang C, Fan C. Attention-based interpretable neural network for building cooling load prediction. *Appl Energy* 2021;299:117238.
- [38] Gao Y, Miyata S, Akashi Y. Multi-step solar irradiation prediction based on weather forecast and generative deep learning model. *Renew Energy* 2022;188:637–50.
- [39] Pourdaryaei A, Mohammadi M, Mubarak H, Abdellatif A, Karimi M, Gryazina E, et al. A new framework for electricity price forecasting via multi-head self-attention and CNN-based techniques in the competitive electricity market. *Expert Syst Appl* 2024;235:121207.
- [40] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9(8):1735–80. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [41] Schuster M, Paliwal K. Bidirectional recurrent neural networks. *IEEE Trans Signal Process* 1997;45(11):2673–81. <http://dx.doi.org/10.1109/78.650093>.
- [42] Mikolov T, et al. Statistical language models based on neural networks. 2012, Presentation at Google, Mountain View, 2nd April 80 (26).
- [43] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. 2014, arXiv preprint [arXiv:1409.0473](https://arxiv.org/abs/1409.0473).
- [44] Jia P, Cao N, Yang S. Real-time hourly ozone prediction system for yangtze River Delta area using attention based on a sequence to sequence model. *Atmos Environ* 2021;244:117917.
- [45] Luong M-T, Pham H, Manning CD. Effective approaches to attention-based neural machine translation. 2015, arXiv preprint [arXiv:1508.04025](https://arxiv.org/abs/1508.04025).
- [46] Li Z, Xu R, Luo X, Cao X, Du S, Sun H. Short-term photovoltaic power prediction based on modal reconstruction and hybrid deep learning model. *Energy Rep* 2022;8:9919–32.
- [47] Brester C, Kallio-Myers V, Lindfors AV, Kolehmainen M, Niska H. Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations. *Renew Energy* 2023;207:266–74.
- [48] Polasek T, Čádik M. Predicting photovoltaic power production using high-uncertainty weather forecasts. *Appl Energy* 2023;339:120989.
- [49] Hossain MS, Mahmood H. Short-term photovoltaic power forecasting using an LSTM neural network and synthetic weather forecast. *IEEE Access* 2020;8:172524–33.
- [50] Markovics D, Mayer MJ. Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction. *Renew Sustain Energy Rev* 2022;161:112364.
- [51] Sarmas E, Spiliotis E, Stamatopoulos E, Marinakis V, Doukas H. Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models. *Renew Energy* 2023;216:118997.
- [52] Shahriari B, Swersky K, Wang Z, Adams RP, De Freitas N. Taking the human out of the loop: A review of Bayesian optimization. *Proc IEEE* 2015;104(1):148–75.
- [53] Zhang W, Wu C, Zhong H, Li Y, Wang L. Prediction of undrained shear strength using extreme gradient boosting and random forest based on Bayesian optimization. *Geosci Front* 2021;12(1):469–77.
- [54] Wang Y, Kandeal A, Swidan A, Sharshir SW, Abdelaziz GB, Halim M, et al. Prediction of tubular solar still performance by machine learning integrated with Bayesian optimization algorithm. *Appl Therm Eng* 2021;184:116233.