A Spatial-temporal Neural Network for Photovoltaic Power Prediction

Yan Huang
Department of Automation
Tsinghua University
Beijing, China
huang_yan@tsinghua.edu.cn

Junwei Cao
Beijing National Research Center of Information Science and Technology
Tsinghua University
Beijing, China
jcao@tsinghua.edu.cn

Shixia Cai
Global Technical Services Business Unit
Alibaba Cloud Intelligence Group
Beijing, China
shixia.csx@alibaba-inc.com

***Abstract*—The access of photovoltaic (PV) systems in the distribution networks has facilitated accurate PV power prediction for energy coordination and operations planning. Due to the power quality and photoelectric consumption of distribution system, the prediction task is still challenging. This paper aims to provide a spatial-temporal network to capture appropriate dependencies between temporal historical PV data and spatial effect factors for the multiple-step situational predictions. It involves a simplified version of convolution and pooling operation used to capture local features and model spatial correlations in the input data, and an attention based bidirectional long short-term memory (BiLSTM) used to enhance that association data prediction. Experimental results demonstrate that the proposed prediction network indicates enhanced accuracy across various time intervals and exhibits superior performance in prediction stability and robustness when compared to typical prediction methods based on the actual measurements from a photovoltaic micro plant. The method excels in capturing both temporal and spatial dependencies, making it a valuable tool for energy planning and operations in distribution networks.**

Keywords—PV power prediction, convolution, BiLSTM, spatial-temporal feature

I. INTRODUCTION

With the intensification of energy crisis and the increase of environmental pollution, there is an increasing emphasis on renewable energy sources. Though PV power holds great potential as a renewable energy technology, its intermittent peculiarity poses a significant challenge for grid integration. Fluctuations in PV power output arise from diverse factors, including solar radiation intensity, shading and obstructions, atmospheric variations and limited PV modules, which make it difficult for the utility grid to maintain a balance between power demand and supply [1]. Consequently, a pivotal and effective measure towards addressing the challenge is to explore an accurate prediction method for PV power generation.

PV prediction refers to the process of estimating the future power output of PV systems based on various factors and data inputs. Accurate and reliable PV prediction plays a crucial role in ensuring proper energy management, grid integration, and operational planning [2]. State-of-the-art works involve various solutions, ranging from physical models [3-7], statistical methods [8-10], machine learning algorithms [11-14], weather forecast integration [15-18], to hybrid approaches [19-22]. Physical models usually use mathematical equations and physical principles to model the behavior of PV systems. They consider factors such as solar radiation, temperature, and panel characteristics to estimate the power output. Typical examples include that single-diode based model [3-5] and the Sandia array performance model [6, 7]. Statistical methods utilize historical data to identify patterns and correlations between PV inputs and output power. For example, regression techniques like multiple linear regression or support vector regression, are often employed to develop prediction models. Machine learning algorithms, such as artificial neural networks, random forests, and support vector machines, are used to train models based on historical data. These models capture complex relationships between input variables and output power, enabling accurate PV prediction [12]. On the other hand, PV prediction can be enhanced by integrating weather forecasts to estimate solar radiation and temperature for future time intervals [23]. They generally allows for more real-time adjustments to expected PV power output. Besides, hybrid methods tend to combine multiple prediction techniques to exploit their respective strengths to achieve general PV prediction tasks. For example, a hybrid model might incorporate physical models, statistical analysis, and machine learning algorithms to improve the prediction accuracy of PV power [24, 25].

PV power prediction requires not only forecasting the future values that are as close to the truth as possible but also maintaining the stability at each step of the prediction in a long time series. As different network structure illustrates different sensitivity and ability to the sequential data, this paper combines the strengths of convolution process and an attention based bidirectional long short-term memory to provide a robust hybrid prediction method. It extracts the spatial and temporal features from the multi-input data and effectively learns from bidirectional sequence information that connected with the *Bahdanau* attention [26]. The goal of this paper is to make the method reliably and accurately predict the PV power data.

II. METHOD FOR PV POWER PREDICTION

A. Overview

The proposed spatial-temporal prediction method is mainly designed upon a simplified version of convolutional neural network module, a bidirectional long short-term memory layer and an attention layer. Fig. 1 demonstrates the framework of the prediction network, which contains the following steps:

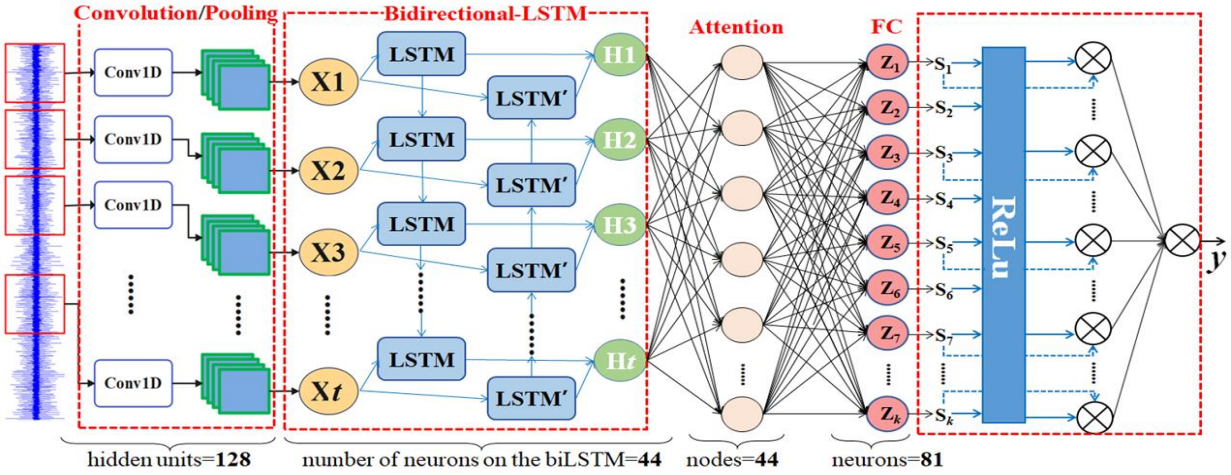

Fig. 1. The framework of the proposed spatial-temporal network for general PV prediction tasks, which depicts four compact connected layers.

1) Preprocessing the input PV power data.
2) Use one-dimensional convolution to extract the spatial features of the data obtained from 1). Select convolution kernels of different sizes to capture feature information of different scales, and use the activation function *ReLU* to process the convolution results nonlinearly.
3) Apply max pooling to conduct the downsampling to reduce the size of the feature map obtained from step 2) and retain the key features.
4) Feature maps from 3) are input into a BiLSTM layer, which learns contextual representations of sequences and generates sequence-level feature representations.
5) The output of step 4) are weighted and integrated in the attention layer to assign different attention weights for different input positions, which makes the network pay more attention to the sequence fragments that have important influence on the multi-step prediction results, and improves the performance and generalization ability of the network.
6) A fully connected (FC) layer is applied to transform the learned features and representations from 5) into the final *ReLU* layer to get the final output.

B. Data Preprocessing

Data preprocessing aims to prepare the input data before it can be fed into the network. In this paper, the preprocessing involves normalization, feature scaling, and missing values handling. Min-Max normalization, as shown in (1), is used to scale the input PV data within a specific range to avoid bias towards certain variables.

$$X'_{norm}(t)=\frac{X_t-X_{min}}{X_{max}-X_{min}}\in[0,1]\,. \tag{1}$$

Normalization by the maximum value of each feature, as shown in (2), is employed to conduct the feature scaling.

$$X'_{scale}(t)=\frac{X_t}{X_{max}}\in[0,1]\,. \tag{2}$$

To handle the missing values in the input data, multiple linear regression strategy, shown in (3), is utilized to estimate missing values based on other features.

$$y(n)=b_0+b_1X_1+b_2X_2+\cdots+b_nX_t\,. \tag{3}$$

Among the three formulas, X_t is the original value, X_{max} is the maximum value of the feature, X_{min} is the minimum one, $X'_{norm}(t)$ represents the normalized value, $X'_{scale}(t)$ represents the scaled value, $y(n)$ represents the target variable (missing value), and $b_0, b_1, b_2, \cdots, b_n$ represent the coefficients or weights associated with each feature.

C. Convolution and Pooling Layer

The one-dimensional convolutional layer applies an array of adaptable filters to slide across the data sequence to perform local feature extraction. Each filter convolves with the local regions of the input, resulting in a sequence of feature maps. By utilizing specific kernel sizes, the convolutional layer can capture different spatial patterns within the data. Each feature map encodes specific patterns or spatial information presented within the input sequence. The convolution layer and linear process are defined with the following (4) and (5), respectively,

$$\mathbf{C}_t=ReLU\left(\sum_{k=1}^{k}W_t\times X_t+b_t\right), \tag{4}$$

$$\mathbf{L}_l=ReLU\left(\mathbf{L}_{l-1}W_l+b_l\right), \tag{5}$$

where $\mathbf{C}_t$ is the resultant feature map subsequent to the intricate mapping process, k denotes the index value of the filter, $\times$ is the convolution operation. X_t is the input sequential data, W_t is donates the weight matrix of the convolution kernel and b_t is represents the bias vector. In (4), $\mathbf{L}_l$ is the output of l-th layer, W_l is the connected weight matrix and b_l is the bias.

The Max pooling operation is applied in the pooling layer to further enhance the spatial feature extraction by downsampling the feature maps. It selects the maximum value within each pooling window to represent the most salient feature present in that region, which reduces the spatial dimensions while retaining the important information and capturing the most relevant features. The feature maps in the pooling layer can be represented with the following equation.

$$\boldsymbol{P}_k=\max_{\rho(n-1)<n<\rho r}\left(X_t(n)\right),\ r=1,2,\ldots\,, \tag{6}$$

where $\boldsymbol{P}_k$ denotes the feature map of the *k-th* filter obtained from that convolution layer, $X_t(n)$ indicates the *n-th* neuron value in the corresponding filter, ρ and r represent the window size and the *r-th* moving step, respectively.

D. Bidirectional Long Short-term Memory Layer

BiLSTM integrates forward and backward computations for sequential data analysis. Its purpose is to capture and model the intricate dependencies and patterns inherent in such data by leveraging the power of Long Short-Term Memory (LSTM) units in both temporal directions. When the features are input into the layer, each structural direction is conducted as follow.

$$\boldsymbol{I}_t = \sigma\left(W_i \mathrm{X}_t + U_i h_{t-1} + b_i\right), \tag{7}$$

$$\boldsymbol{f}_t = \sigma\left(W_f \mathrm{X}_t + U_f h_{t-1} + b_f\right), \tag{8}$$

$$\boldsymbol{o}_t = \sigma\left(W_o \mathrm{X}_t + U_o h_{t-1} + b_o\right), \tag{9}$$

$$\boldsymbol{g}_t = tanh\left(W_g \mathrm{X}_t + U_g h_{t-1} + b_g\right), \tag{10}$$

$$\boldsymbol{ce}_t = \boldsymbol{f}_t \odot \boldsymbol{ce}_{t-1} + \boldsymbol{I}_t \odot \boldsymbol{g}_t, \tag{11}$$

$$\boldsymbol{b}_t = \boldsymbol{o}_t \odot tanh\left(\boldsymbol{ce}_t\right), \tag{12}$$

where $\boldsymbol{I}_t$ denotes the input gate, $\boldsymbol{f}_t$ is the forget gate, $\boldsymbol{o}_t$ is the output gate, $\boldsymbol{g}_t$ is the candidate value, $\boldsymbol{ce}_t$ is the cell state, and $\boldsymbol{b}_t$ is the hidden state, X_t represents the input at time step t, h_{t-1} denotes the previous hidden state, W and U are weight matrices, b is the bias term, σ denotes the sigmoid activation function, *tanh* is the hyperbolic tangent activation function, and $\odot$ is element-wise multiplication. Forget gate determines the extent by which the previous memory content should be forgotten. Input gate controls the amount of new information to be stored in the memory. The candidate value represents the potential new information to be added to the memory. The output gate regulates the information to be passed from the memory to the hidden state. The cell state represents the updated memory at time step t. And the hidden state represents the output or the information propagated to the next time step.

E. Bahdanau Attention Layer

BiLSTM has shown effectiveness in learning from shorter input sequences. However, when applied to longer sequences, it will suffer from information loss, which makes it challenging to obtain a meaningful vector representation of the entire input sequence. To address this limitation, we use that *Bahdanau* attention [26] to learn the significance of each element in the sequential data and assign higher attention to crucial parts that have a significant impact on the output results. Fig. 2 depicts the workflow of the attention mechanism on an input data sequence (x_1, x_2, …, x_t) to generate the t-th target value y_t, where H_1 to H_t represents the forward or backward hidden states at time t. By effectively leveraging the mechanism, the network adeptly amalgamates both forward and backward information, which empowers it to make informed predictions and proficiently handle longer input sequences.

F. Fully Connected & ReLU Layer

In the proposed network, we design a FC layer to perform a linear transformation on the input data, mapping it to the output space. As in Fig. 2, the output of the attention layer is denoted as y_t (t=0, 1, 2…) with a size of n. The weight matrix of the FC layer is represented by W^T, with a shape of (m, n), where m represents the number of neurons in the layer. Additionally, there is a bias vector b^T, with a size of (m, 1). As shown in Fig. 1, the calculation formula for the output of fully connected layer

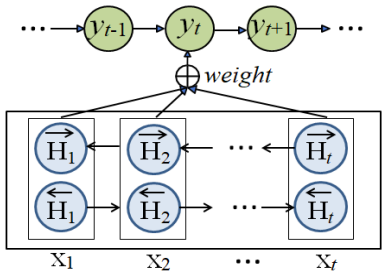

Fig. 2. The workflow of *Bahdanau* attention in our proposed prediction network.

can be expressed as equation (13).

$$\mathrm{Z}_k = \boldsymbol{W}_X + b^T, \tag{13}$$

where W_X denotes the matrix multiplication between the weight matrix W^T and the input X_t, and Z_k represents the output of the fully connected layer.

On the other hand, the output of the FC layer is designed to further pass through a non-linear activation function *ReLU* to increase the expressiveness of the network. It introduces non-linearity to the calculation formula, enabling more complex patterns and relationships to be captured.

The FC layer is used to act as a mapping layer to transform the features or representations from the previous layer into the final output. Through a series of linear transformations and non-linear activations, the FC and *ReLU* layers facilely capture the intricate patterns and relationships in the input data, which facilitates the generation of the desired output for the network.

G. Training

Fig. 3 illustrates the training flow of the proposed PV prediction network. It involves the following several key steps.

*Step*1: Initialize all the weight parameters of the neurons.

*Step*2: Perform forward computation to calculate the output values of each neuron. In the BiLSTM layer, it includes the vector values of $\boldsymbol{f}_t$, $\boldsymbol{I}_t$, $\boldsymbol{o}_t$, $\boldsymbol{ce}_t$, and h_t.

*Step*3: Compute the error term δ for each neuron. Propagate the error backward in time, starting from the current t time step and calculating the error term for each time step, and then propagate the error term to the previous layer.

*Step*4: Calculate the gradients of each weight parameter based on the corresponding error term δ.

*Step*5: Utilize the optimization strategy *Adam* to update the weight parameters, and continue iterating until the number of iterations reaches the maximum or the total error converges to the specified threshold.

III. ANALYSIS OF EXPERIMENTAL RESULTS

A. Experimental Design

In order to verify the prediction accuracy of the proposed network, this paper construct five different related networks with varying complexities for evaluation, including CNN, LSTM, BiLSTM, CNN-LSTM, LSTM-attention and BiLSTM-attention. We use the actual measurements from a photovoltaic micro plant with resolution of 5 minutes per point to perform the experimental comparisons by conducting future multi-step prediction. Multi-step prediction in this paper means that the

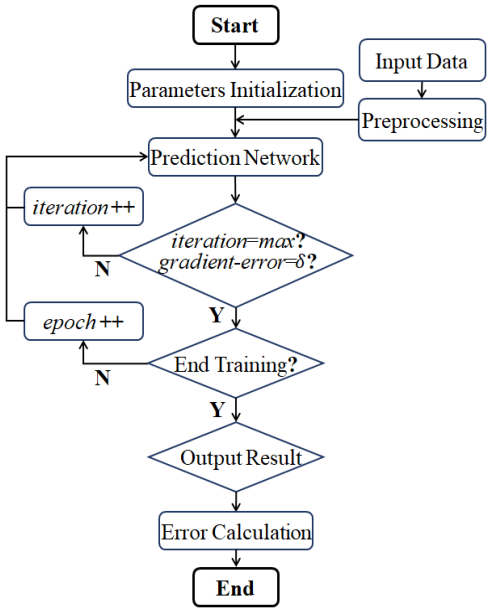

Fig. 3. The training process of the proposed PV prediction network.

networks need to forecast all the values from the next point in time to 24 o'clock on the same day at one time. The metrics of mean absolute percentage error (MAPE), root mean square error (RMSE), mean absolute error (MAE), and R-squared (R^2) are applied to evaluate the quality of the comparative networks. Lower MAPE, RMSE and MAE values indicate a higher accuracy and reliability of the prediction network. A higher R^2 value implies a better fit between the predicted and actual values. The collaboration of these metrics can elaborately describe the integral performance of a network.

B. Subjective Evaluation

Fig. 4 describes some sampling points of the PV power prediction results from the comparative models, which is plotted by overlaying the predicted PV power output values from the model with the actual measured values. It indicates that the predicted values of the seven comparison networks have good overlap with the trend of the actual values. From the sampling results, the proposed network presents a prediction result closer to the actual value, which can be further verified in Fig. 5. It presents future 12 days perdition data, whose results accurately predict the occurrence of peak and off-peak periods, as well as the timing of inflection points, which successfully captures the highs and lows of PV generation, enabling accurate predictions of when the system will reach its maximum and minimum output. From Fig. 5, the proposed model effectively identifies the turning points, indicating the moments when the PV output transitions from increasing to decreasing or vice versa. This indicates that the network is able to capture the underlying patterns and variations in the PV power output accurately. The alignment between the curves suggests that the network's predictions are in line with the actual PV power production, thereby validating the effectiveness of the forecasting model.

C. Objective Evaluation

Table 1 illustrates the evaluated metrics of different prediction methods. From the table, we can conclude that our network achieves obvious performance on the data of metrics.

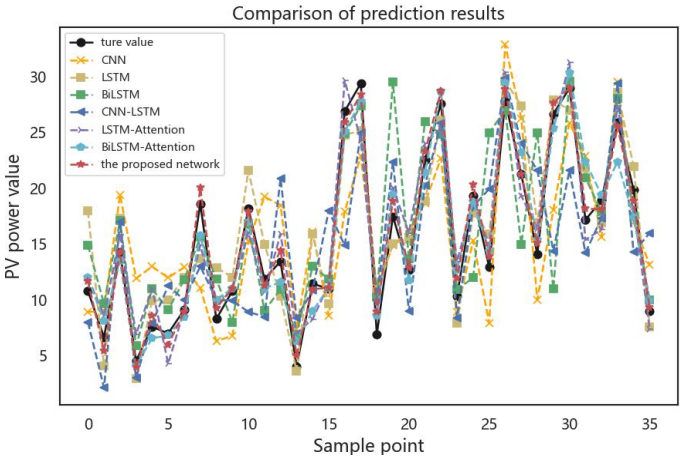

Fig. 4. The PV power prediction results of the seven comparative networks.

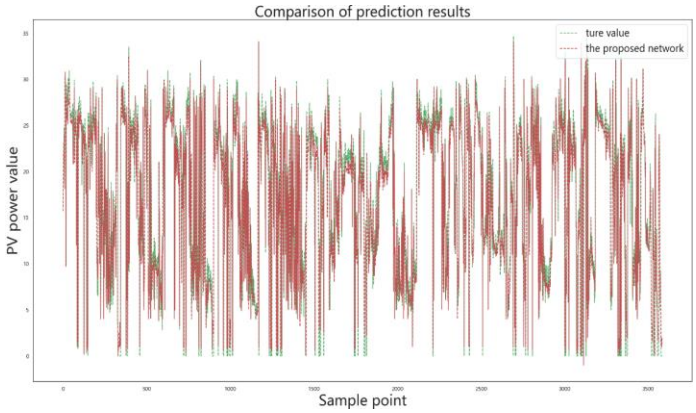

Fig. 5. Multi-step prediction results of ten days from the proposed network.

TABLE I. THE EVALUATED METRICS OF DIFFERENT PREDICTION METHODS.

Method	Average Value of Each Metric			
	MAPE	**RMSE**	**MAE**	**R^2**
CNN	0.071	31.019	14.662	0.713
LSTM	0.064	30.356	14.108	0.785
BiLSTM	0.041	30.140	12.135	0.806
CNN-LSTM	0.039	26.006	10.805	0.911
LSTM-Attention	0.036	19.788	8.693	0.934
BiLSTM-Attention	0.032	16.719	8.030	0.957
The proposed network	**0.021**	**9.083**	**5.992**	**0.991**

Specifically, It reduces 70.42% in MAPE , 70.72% in RMSE, 59.13% in MAE and gains 38.99% improvement in R^2 on average when compared with that CNN prediction network. The proposed method reduces 67.19% on average in MAPE, 70.08% in RMSE, 57.53 % in MAE and gains 26.24% improvement in R^2 over LSTM network, and reduces 48.78% on average in MAPE, 69.86% in RMSE, 50.62 % in MAE and gains 22.95% improvement in R^2 over BiLSTM network, and it can reduce 46.15 % on average in MAPE, 65.07% in RMSE, 44.55 % in MAE and gains 8.78% improvement in R^2 over CNN-LSTM, reduces 41.67% on average in MAPE, 54.10% in that RMSE, 31.07 % in MAE and gains 6.10% improvement in R^2 over LSTM-Attention network, and reduces 34.38% on average in

MAPE, 45.67% in RMSE, 25.38 % in MAE and gains 3.55% improvement in R^2 over BiLSTM-Attention. The significant lower MAPE, RMSE, MAE values and the higher R^2 clearly establish the superior accuracy, precision, and reliability of the proposed network when compared to other prediction networks.

IV. Conclusion

In this paper, we propose a novel spatial-temporal network for general photovoltaic power prediction tasks. The motivation behind the method is that a convolution and pooling layer is responsible for extracting spatial features from the multi-dimensional data, a BiLSTM layer tends to capture long-term temporal dependencies by learning and memorizing sequential context information progressively through multiple sequential steps, and an attention is employed to weight and integrate the outputs of the two layers to pursue enhanced prediction accuracy. In comparison to comparative perdition models, the proposed network yields more effective performance in terms of improved accuracy and robustness in PV power prediction. The results indicate its great potential on predicting various challenging PV power data.

References

[1] J. Liu, H. Wang, and T. Hao, "Short-term photovoltaic power prediction based on bayesian regularized bp neural networks," International Conference on Electrical Engineering and Green Energy (CEEGE), Grimstad, Norway, 2023, pp. 224-229.

[2] Y. -J. Liu, Y. -D. Lee, C. -Y. Lee, C. -C. Cheng, P. -Y. Hou, and Y.-F. Chen, "A comparative analysis of lstm and bilstm network-based methods in pv power prediction," IET International Conference on Engineering Technologies and Applications (IET-ICETA), Changhua, Taiwan, 2022, pp. 1-2.

[3] H. Shi and C. Xu, "Photovoltaic power prediction based on a modified Single-Diode model using particle swarm optimization," Energies, vol. 11, no. 9, pp. 2272, 2018.

[4] P. J. J. Botha, and A. J. Rix, "Correcting the induced errors when using De Soto's parameters in the PV single diode model," IEEE AFRICON, Cape Town, South Africa, 2017, pp. 1113-1118.

[5] P. Chen, J. Liu, and W. Cao, "Photovoltaic power forecasting using a Single-Diode model and support vector regression," Energies, vol. 11, no. 6, pp. 1499, 2018.

[6] W. Yuan, H. Qin, Y. Zhang, and Z. Luo, "Photovoltaic power forecasting based on an improved Sandia Array Performance Model," Applied Energy, vol. 242, pp. 874-883, 2019.

[7] R. Azizipanah-Abarghooee, M. Hejazy, M. R. Akbarzadeh-T, S. M. Mirnezhad, E. Charkhgard, B. Tork Ladani, A. Arabkoohsar, and T. Niknam, "Short-term PV power forecasting using an optimized ARMAX model with hourly global solar radiation," Solar Energy, vol. 226, 2021.

[8] T. Ma, H. Yang, and Y. Li, "Statistical modeling and forecast of photovoltaic power based on WT-ARMA-GARCH," Applied Energy, vol. 238, pp. 721-731, 2019.

[9] X. Qiao, Y. Wang, and M. Xu, "A statistical forecasting method for photovoltaic power output based on a seasonal adaptive decomposition ensemble model," Renewable Energy, vol. 147, pp. 182-198, 2020.

[10] B. Wang, X. Deng, T. Chen, S. Zheng, and D. Li, "Distributed Photovoltaic Daily Power Generation Forecast Method Based on Model Selection and Statistical Scale-up," International Conference on Electronics Technology (ICET), Chengdu, China, 2023, pp. 842-847.

[11] C. Li, F. Wen, B. Zhang, and Y. Zhang, "An accurate photovoltaic power prediction method based on biased support vector regression optimized by krill herd algorithm and bat algorithm," Applied Energy, vol. 284, pp. 116553, 2021.

[12] R. P. Ponraj, V. Ravindran, K. Chittibabu, A. T, C. A, and E. Louis, "Machine learning based solar pv power prediction," International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, 2023, pp. 1646-1652.

[13] S. Kumar.A, G. Pv, and B. Jackson, "Machine learning-based timeseries analysis for cryptocurrency price prediction: a systematic review and research," International Conference on Networking and Communications (ICNWC), Chennai, India, 2023, pp. 1-5.

[14] R. Zheng, G. Li, K. Wang, B. Han, Z. Chen, and M. Li, "Short-term Photovoltaic Power Prediction Based on Daily Feature Matrix and Deep Neural Network," Asia Conference on Power and Electrical Engineering (ACPEE), Chongqing, China, 2021, pp. 290-294.

[15] S. Tiwari, R. Sabzehgar, and M. Rasouli, "Short Term Solar Irradiance Forecast Using Numerical Weather Prediction (NWP) with Gradient Boost Regression," IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG), Charlotte, NC, USA, 2018, pp. 1-8.

[16] C. Poolla, and A. K. Ishihara, "Localized solar power prediction based on weather data from local history and global forecasts," IEEE World Conference on Photovoltaic Energy Conversion (WCPEC), Waikoloa, HI, USA, 2018, pp. 2341-2345.

[17] B. Huo, J. Zhang, and C. Lan, "Photovoltaic power forecasting based on a hybrid model integrating weather forecast data and artificial neural networks," Energies, vol. 14, no. 11, pp. 3303, 2021.

[18] B. Li, C. Chen, and X. Zhu, "Photovoltaic power prediction combining a theoretical model and weather forecast data using LSTM network," IEEE Transactions on Sustainable Energy, vol. 12, no. 3, pp. 2253-2263, 2021.

[19] J. Shi, Y. Chen, X. Cheng, M. Yang, and M. Wang, "Four-stage space-time hybrid model for distributed photovoltaic power forecasting," IEEE Transactions on Industry Applications, vol. 59, no. 1, pp. 1129-1138, Jan.-Feb. 2023.

[20] W. Lin, B. Zhang, and R. Lu, "Short-term forecasting for photovoltaic power based on successive variational modal decomposition and cascaded deep learning model," International Conference on Energy, Electrical and Power Engineering (CEEPE), Guangzhou, China, 2023, pp. 1011-1016.

[21] L. Stoyanov and I. Draganovska, "Comparison of hybrid models for pv power output forecasting—application to Oryahovo, Bulgaria," Conference on Electrical Machines, Drives and Power Systems (ELMA), Varna, Bulgaria, 2023, pp. 1-4.

[22] P. Zou, F. Liu, J. Wang, and L. Zhu, "Hybrid photovoltaic power forecasting model by combining empirical mode decomposition and gated recurrent unit," Applied Energy, vol. 281, pp. 116065, 2021.

[23] C. Poolla, and A. K. Ishihara, "Localized solar power prediction based on weather data from local history and global forecasts," IEEE World Conference on Photovoltaic Energy Conversion (WCPEC), Waikoloa, HI, USA, 2018, pp. 2341-2345.

[24] J. Yang, T. Wu, K. Wang, and R. Wen, "A hybrid vmd-based ARIMA-LSTM model for day-ahead pv prediction and uncertainty analysis," International Conference on Smart Power & Internet Energy Systems (SPIES), Beijing, China, 2022, pp. 2009-2014.

[25] W. Han, Z. Tang, Z. Xu and M. Chen, "Hybrid Model Based on EEMD, ARMA and Elman for Photovoltaic Power Prediction," International Conference on Intelligent Control, Measurement and Signal Processing (ICMSP), Hangzhou, China, 2022, pp. 438-441.

[26] D. Bahdanau, K. Cho, Y. Bengio. "Neural machine translation by jointly learning to align and translate," Computer Science. 2014, http://arxiv.org/abs/1409.0473v6.