PIER C

(Received 13 February 2025, Accepted 1 April 2025, Scheduled 13 April 2025)

Photovoltaic Power Prediction Based on K-Means++ -BiLSTM-Transformer

Jianwei Liang[1], Liying Yin[1], Sichao Li[1], Xiubin Zhu[1], Zhangshen Liu[1], and Yanli Xin[2, *]

[1] *School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, Ganzhou, Jiangxi 341000, China*
[2] *School of Automattion, Guangdong Polytechnic Normal University, Guangzhou, Guangdong 510665, China*

ABSTRACT: The inherent volatility and uncertainty associated with photovoltaic (PV) power generation present significant challenges to maintaining grid stability. As the level of PV integration into the grid continues to rise, the importance of accurately predicting its power output becomes increasingly critical. This study presents a new PV power prediction model utilizing K-means++-Bidirectional Long Short-Term Memory (BiLSTM)-Transformer framework. Initially, the Pearson correlation coefficient is computed to determine the key factors influencing the prediction of PV power significantly. Following this, the K-means++ clustering algorithm is applied to analyze historical power data, categorizing it into three distinct groups corresponding to different weather conditions. Finally, the BiLSTM-Transformer architecture is employed to develop a power output prediction model tailored for the three weather scenarios. The prediction model is subsequently optimized using Bayesian methods to determine the optimal model configuration for each specific weather condition. Experimental findings demonstrate that the proposed K-means++-BiLSTM-Transformer similar day PV power prediction model exhibits superior accuracy, enhanced generalization, and increased robustness compared to alternative prediction models.

1. INTRODUCTION

Over the past few years, solar power generation, characterized by its clean and renewable nature, has progressively displaced traditional thermal power generation within the energy sector and has achieved extensive application [1]. However, the rapid growth of PV installed capacity poses challenges. The large-scale integration of PV power into the electrical grid has become a notable challenge to the stability of the grid system. Moreover, the intermittent and highly variable characteristics of PV power further complicate management efforts [2, 3]. Thus, accurate prediction of PV output is essential for ensuring grid stability and improving the efficiency of energy management systems. By increasing the reliability of solar power generation forecasts, operators can better balance supply and demand, thereby enhancing overall grid performance. Improved forecasting accuracy directly contributes to more effective resource allocation and supports the development of robust energy management frameworks in modern power networks.

Contemporary PV forecasting methods fall into three main categories: physical models, statistical techniques, and artificial intelligence (AI)-based approaches. Physical methods simulate solar conversion processes, and statistical tools analyze historical data patterns, while AI-based solutions employ advanced machine learning algorithms [4–8]. Ref. [9] suggests utilizing temporal patterns in PV system operations to enhance output forecasting accuracy. The approach focuses on analyzing sequential data to capture solar generation dynamics, offering improved predictive performance through time-series analysis. However, the physical modeling process is complex, and the prediction effect is not good under extreme weather conditions. Although the autoregressive method was used in [10] and [11] to predict PV power, the statistical method relied on a large amount of historical data and had limited ability to extract non-stationary information in time series. In contrast, AI-based algorithms can dig deep into time series information, and the model has strong self-learning capabilities, showing significant advantages.

Advances in machine learning and deep learning have greatly enhanced PV power forecasting, with AI-driven methods achieving significant success. These approaches have markedly improved prediction accuracy, proving their effectiveness in solar energy generation forecasting. Lee and Kim [12] proposed a Recurrent Neural Network (RNN)-based framework adept at capturing long-term dependencies in sequential data, improving pattern recognition across extended time intervals in PV generation records. The model also mitigates the vanishing gradient problem, enhancing training stability. However, it overlooks key weather factors such as solar irradiance and cloud cover, potentially reducing forecast accuracy and practical applicability. Ref. [13] proposed a PV power prediction model combining improved C-mean clustering with Long Short-Term Memory (LSTM), achieving better historical sample classification and prediction accuracy. However, the model lacks hyperparameter optimization, which is essential for performance enhancement. In [14], a power prediction model integrating Convolutional Neural Network (CNN), BiLSTM, and attention mechanisms was developed to minimize the loss of key historical information. Tests on actual power data showed significant accuracy improvements, though the model did not account for weather variations, leading to

* Corresponding author: Yanli Xin (yanlixin@gpnu.edu.cn).

potential performance fluctuations under different meteorological conditions. Ref. [15] presented a Transformer-based PV power prediction model, analyzing factors like irradiation intensity and PV array direction to build predictive models. Multi-site experimental validation confirmed its ability to capture long-term temporal dependencies, demonstrating consistent performance across diverse locations. However, the modeling process did not differentiate between various weather conditions, which may lead to variations in model performance across different meteorological scenarios. In this context, [16] introduced a hybrid model combining optimized K-means clustering, gray relational analysis, and Elman networks, offering enhanced predictive capabilities through integrated analytical approaches. The optimized K-means method clusters historical power data, identifying days with patterns similar to those of the forecast target. This approach enhances prediction accuracy through relevant historical sample selection. The power and meteorological data from these similar days were then fed into the model for training purposes. Compared with several other prediction methods, this approach demonstrated significantly reduced errors and higher prediction accuracy. Given that PV power prediction involves long time series forecasting, the prediction model must possess a strong capability to capture information within the time series. According to the literature [17–19], K-means++ demonstrates certain advantages over several common clustering methods. For instance, compared to Fuzzy c-means clustering, K-means++ has lower computational complexity and produces more distinct clustering results. In comparison to hierarchical clustering, K-means++ is more computationally efficient, and its algorithmic logic is relatively simple, making it easier to implement. Compared to the standard K-means clustering, K-means++ offers better initial centroid selection, resulting in more stable outcomes and faster convergence.

This study presents a novel forecasting framework that integrates K-means++ clustering, BiLSTM networks, and transformer architecture to address limitations in PV power prediction. During preprocessing, Pearson correlation analysis identifies key meteorological factors linked to PV output, followed by K-means++ clustering of historical data. For each weather type, a BiLSTM-Transformer model is developed for predictions. Bayesian optimization is applied to fine-tune critical parameters, significantly enhancing model performance. This method systematically identifies optimal configurations, improving the predictive power of the combined BiLSTM-Transformer structure. The tuning process ensures more accurate and reliable forecasts through optimized parameterization. BiLSTM networks excel at handling long sequences by capturing both past and future information, enriching feature extraction. The Transformer model complements this by efficiently capturing global sequence characteristics. Its multi-head attention mechanism extracts features across different positions and performs parallel computations in multiple subspaces, enabling a comprehensive representation of complex temporal relationships. This ensures robust feature learning across diverse time scales and patterns in the input data.

2. RESEARCH METHODS

2.1. Pearson Correlation Coefficient

Pearson correlation coefficient serves as a quantitative indicator for assessing linear relationships between paired variables, with its numerical values spanning from perfect negative correlation (-1) to perfect positive correlation (1). An explanation of the numerical range and correlation of this coefficient is provided in Appendix A. The mathematical representation of this statistical measure is expressed through the following equation.

$$\rho_{x,y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \tag{1}$$

where $\rho_{x,y}$ quantifies the degree of linear correlation between the meteorological variable X and generated power Y, where E denotes the mathematical expectation.

2.2. K-Means++

K-means++ represents an advanced iteration of the conventional K-means clustering method, primarily developed to mitigate the inherent limitations associated with the initialization of cluster centroids in the standard K-means approach. Through an optimized initialization process, K-means++ achieves better clustering performance and reduces the impact of initial centroid placement, thereby yielding more stable and reliable clustering outcomes.

Arthur and Vassilvitskii, in 2007, presented a significant improvement over the conventional K-means method through its refined approach to initial cluster center selection [20]. This optimization ensures that the initial cluster centers are more widely dispersed, thereby improving the overall clustering performance. The implementation of the algorithm consists of the following steps.

(1) A random sample is selected from the dataset to serve as the first cluster center C_1.

(2) Following this, the minimum distance between each data point and the current cluster centers is computed. The likelihood of selecting each sample as the subsequent cluster center is then determined using the following formula.

$$\frac{D(x)^2}{\sum_{x \in X} D(x)} \tag{2}$$

where $D(X)$ denotes the distance from sample x to the closest cluster center.

(3) This selection process is iterated until the desired number of cluster centers K is achieved.

(4) Once the centers are established, each sample is allocated to the nearest cluster center.

(5) Subsequently, the centroid of each cluster is updated by calculating the mean of all samples within that cluster.

$$C_i = \frac{1}{|c_i|} \sum_{x \in c_i} x \tag{3}$$

where c_i represents the set of samples in cluster i.

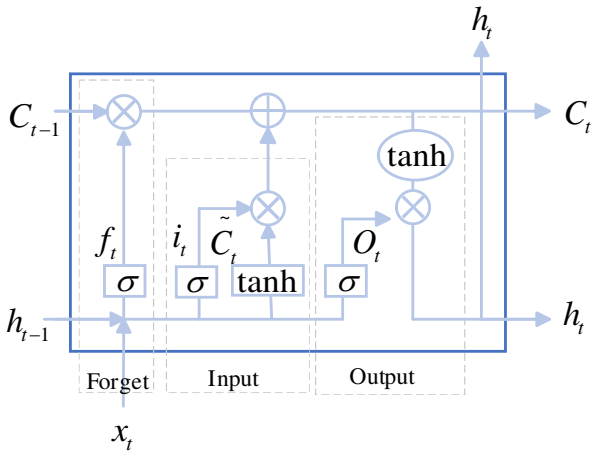

FIGURE 1. LSTM structure.

(6) The final two steps are reassigning samples to the nearest cluster and recalculating cluster centroids, when the cluster centers stabilize. This methodology, known as K-means++ algorithm, optimizes the initialization of cluster centers, thereby improving both the reliability and precision of the clustering outcome.

2.3. BiLSTM Model

The BiLSTM neural network comprises two layers of LSTM units: one layer processes the forward information of the time series, while the other processes the reverse information. LSTM networks, which incorporate memory cells and gating mechanisms, improve traditional RNNs by addressing issues related to short-term memory and gradient explosion [21, 22]. The architecture of an LSTM network includes three sigmoid gating layers (forget gate, input gate, and output gate) as well as a tanh-activated cell state layer, as depicted in Figure 1. However, the unidirectional transmission of parameters in LSTM networks limits their ability to capture deep features within time series data. In contrast, BiLSTM enhances feature extraction by integrating both forward and reverse LSTM layers, allowing for a more comprehensive analysis of temporal relationships and maximizing the utilization of information from different time points. The structural design of BiLSTM is illustrated in Figure 2, and its fundamental equations have been provided. In short, BiLSTM enhances the ability to capture time series features through the combination of forward and reverse LSTM layers.

In Figure 1, the forget gate plays a critical role in identifying and eliminating irrelevant information from the memory unit. The mathematical expression for this operation is provided below.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

where f_t represents the forget gate's output; σ denotes the activation function; W_f corresponds to the weight matrix associated with the forget gate; b_f signifies the bias term; h_{t-1} reflects the hidden state from the preceding time step; and x_t indicates the input at the current time step. The input gate is responsible for selecting which new information should be retained within the memory unit. The mathematical formulation

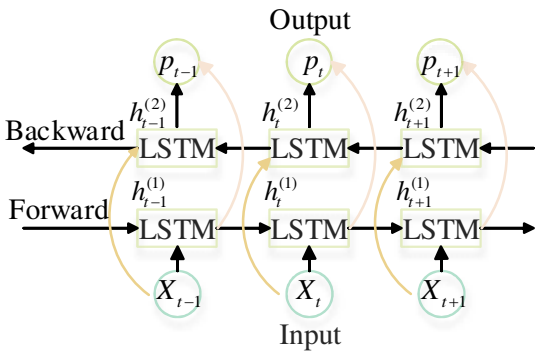

FIGURE 2. BiLSTM structure.

for this process is provided below.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{6}$$

where i_t denotes the activation output of the input gate, and $\tilde{C}$ represents the candidate value of the new information. The weight matrices for the input gate and candidate value are denoted by W_i and W_C, respectively. Additionally, b_i and b_C correspond to the bias terms for the input gate and candidate value, respectively. The updated state of the memory unit is represented by C_t, and the formula is as follows.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{7}$$

where C_{t-1} represents the state of the memory unit at the preceding time step. The output gate is responsible for computing the hidden state h_t, which is derived using the following formula.

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{8}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{9}$$

where O_t denotes the activation output of the output gate, while W_O also represents the weight matrix associated with this gate, and b_O corresponds to the bias term of the output gate.

2.4. Transformer Model

The Transformer is a deep learning architecture designed for sequence-to-sequence tasks, relying solely on attention mechanisms and structured into an encoder-decoder framework [23]. The encoder transforms the input sequence into a high-dimensional vector representation, which is then utilized by the decoder to produce the output sequence. Each layer within the encoder consists of two main components: a self-attention mechanism and a Feed-Forward Network (FFN). These sublayers are connected via layer normalization and residual connections. The decoder shares a comparable architecture with the encoder, but it incorporates an extra attention sublayer within every layer, which is specifically designed to attend to the output generated by the encoder. The Transformer processes sequence data in parallel, achieving high computational efficiency and capturing long-range dependencies through its attention mechanisms. Its structural diagram is illustrated in Figure 3.

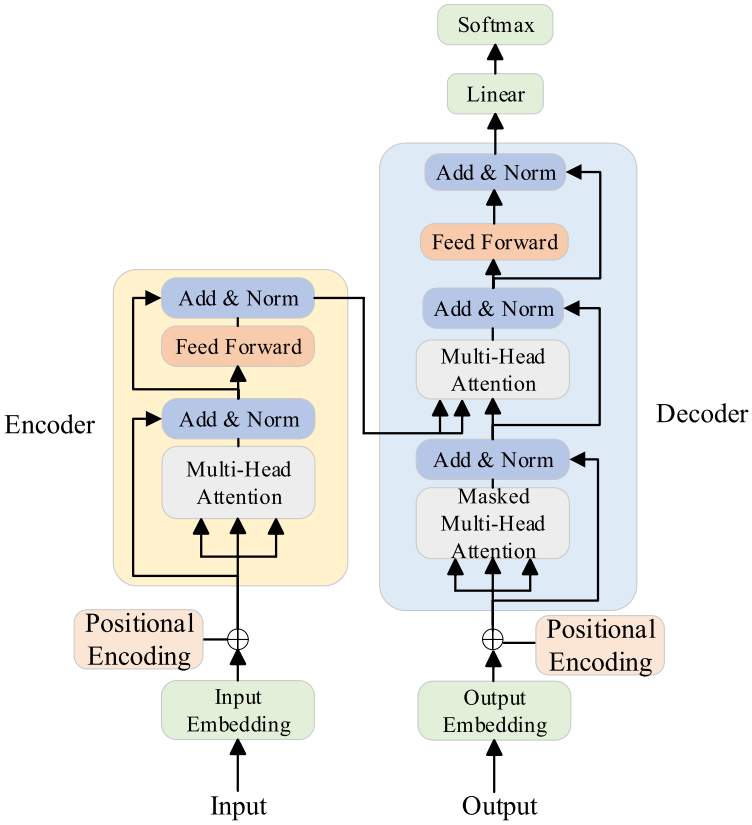

FIGURE 3. Transformer model structure.

Positional Encoding (PE) is the positional relationship between the data in the sequence, and positional encoding needs to be added before the data are input to the multi-head attention layer. In this paper, a sinusoidal function of different frequencies is used to represent the positional relationship between the data.

$$PE(k,2u) = \sin(k/10000^{2u/m}) \quad (10)$$

$$PE(k,2u+1) = \cos(k/10000^{2u/m}) \quad (11)$$

where k represents the time step, u the feature dimension, and m the length of the input sequence. The multi-head self-attention mechanism is employed to model relationships among data points at various positions, and its computation is expressed as follows.

$$MultiHead(Q,K,V) = Concat(head_1,\ldots,head_h)W^O \quad (12)$$

$$head_n = Attention(Q_n,K_n,V_n) \quad (13)$$

$$Attention(Q,K,V) = soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (14)$$

where the multi-head self-attention mechanism is denoted as MultiHead; Q, K, and V correspond to the query, key, and value matrices respectively. These matrices are associated with trainable weight parameters. The variable n signifies the total number of attention heads utilized in the mechanism. Attention represents the attention mechanism. Additionally, d_k, which corresponds to the number of columns in the matrix, is explicitly defined to clarify the structural dimensions involved in the computation.

FFN builds upon the attention mechanism to extract additional features, thereby enhancing the model's expressiveness. The FFN is composed of two linear transformations, which can be expressed as follows.

$$FFN(\delta) = \mathrm{Re}LU(\delta W_1 + b_1)W_2 + b_2 \quad (15)$$

where FFN represents the feedforward network; δ represents the input data; W_1, W_2 correspond to the weight matrix; b_1 and b_2 denote the bias item.

2.5. Bayesian Optimization

The essence of Bayesian optimization algorithms is based on probabilistic surrogate models and sampling functions [24]. The probabilistic surrogate model approximates the objective function by constructing a surrogate function, fitting the objective function with a small number of known points and continuously modifying the prior probability through iteration. The sampling function samples the areas that are not fully explored by the surrogate function and the area where the minimum value is most likely to occur, and gradually updates the surrogate function to bring it closer to the real objective function until the loss function reaches a minimum. The Bayesian hyperparameter optimization steps are as follows:

(1) The initialization point is randomly generated, and the Gaussian process is estimated and updated.

(2) The sampling function is used to guide the new sampling.

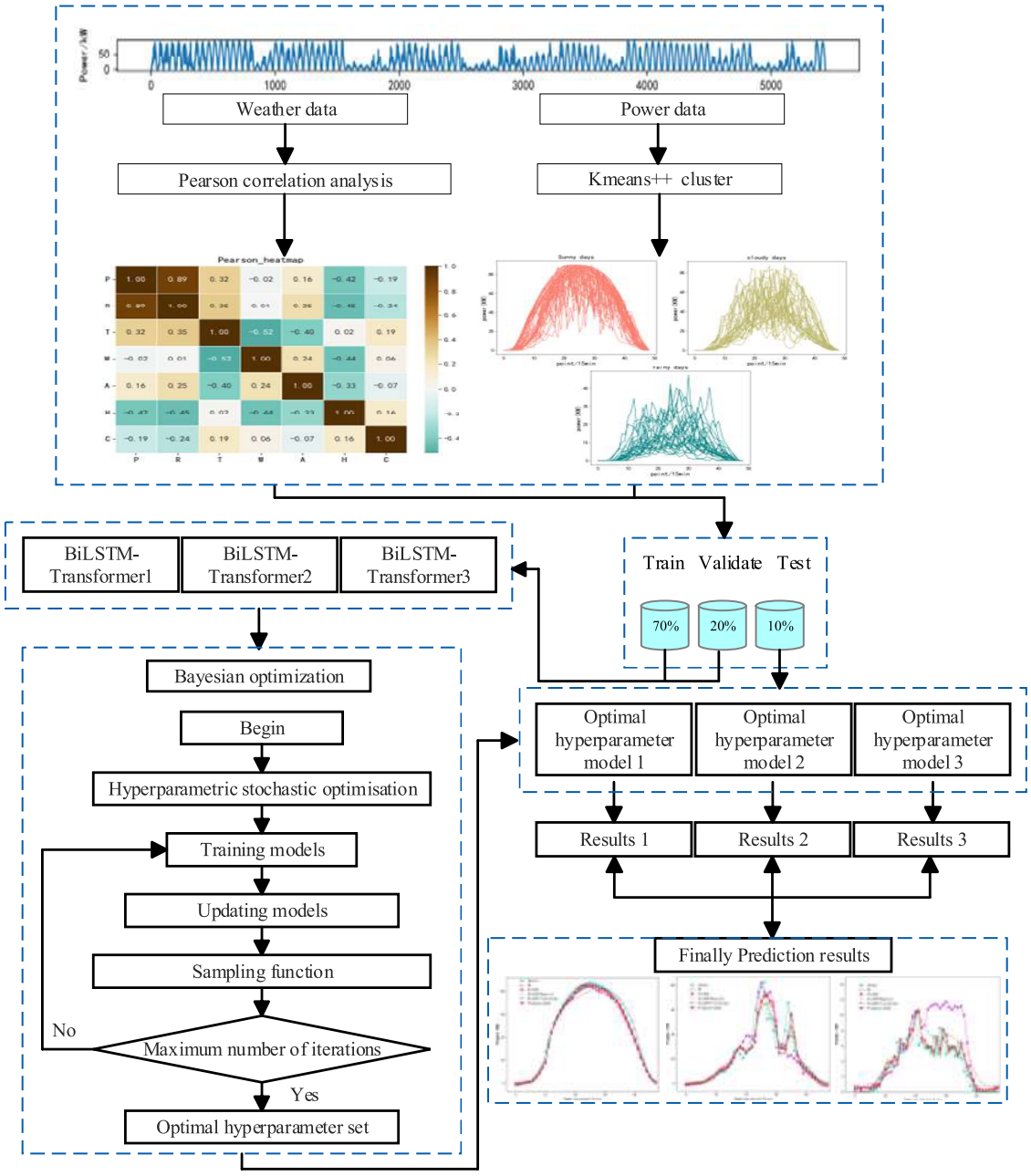

FIGURE 4. K-means++-BiLSTM-Transformer PV power prediction model framework.

(3) The objective function is evaluated on the newly selected sampling point, and the new sampling point and its performance indicators are fed back to the surrogate model. The probabilistic surrogate model is updated.

(4) Repeat steps (2)–(3) to optimize the surrogate model step by step.

(5) The optimization procedure concludes upon reaching the predefined maximum iteration count.

2.6. K-Means++-BiLSTM-Transformer Model

The output of PV systems exhibits significant variability due to meteorological factors, with datasets belonging to identical categories potentially demonstrating markedly different characteristic distributions. When predictive models are used to learn and forecast power generation, interference from irrelevant data may affect the learning process, thereby reducing prediction accuracy. To address this issue, the K-means++ algorithm was employed to cluster the PV power data and develop prediction models tailored to different weather conditions. PV power forecasting represents a specialized domain within time-series analysis, where the effectiveness and precision of forecasting models are fundamentally dependent on the capacity to identify and process temporal characteristics within historical datasets. By employing the BiLSTM network to model historical PV power data, the forward LSTM captures the forward

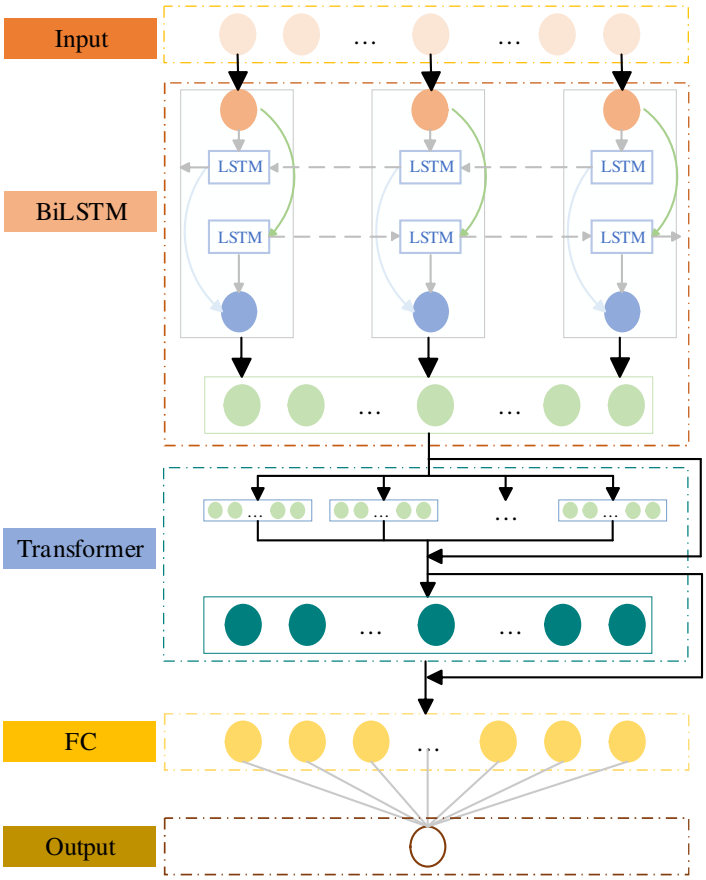

FIGURE 5. BiLSTM-Transformer model structure.

temporal information while the backward LSTM captures the backward temporal information within the time series, thereby enhancing prediction accuracy. Moreover, the Transformer architecture, through its innovative integration of multi-head self-attention mechanisms and FFN networks, demonstrates superior capability in performing comprehensive data analysis and extracting global patterns. This structural design enables the model to establish long-range dependencies and process complex relationships within the data more effectively than traditional approaches. Building upon the previously discussed methodologies, this study introduces a novel approach for PV power forecasting in short-term scenarios, utilizing a hybrid architecture that combines K-means++-BiLSTM-Transformer components. Figure 4 presents the comprehensive architecture of the proposed prediction system, while Figure 5 provides a detailed visualization of the BiLSTM-Transformer network configuration. Within this architecture, the Transformer's encoding module serves as the fundamental building block for model construction.

The proposed model's input features incorporate both atmospheric condition parameters and past PV generation records. The data are formatted as a matrix of size $M \times N$, where M denotes the temporal dimension of the input sequence, and N denotes the information content at each time point within the series. The data with the size of $M \times N$ is input into the BiLSTM network, and the information on different time scales is learned through the bidirectional LSTM network, which effectively captures the relevant information between the data in the long time series. Subsequently, the feature representations generated by the BiLSTM architecture are fed into the Transformer module. The multi-head attention component processes these features to capture positional dependencies across various locations in the sequence. Then, the nonlinear transformation is carried out through the FFN. To facilitate effective information flow and gradient propagation, residual connections are implemented at two critical stages: following the multi-head attention computation and after the FFN processing. In the final stage, the feature representations processed by the Transformer module are fed into the Fully Connected (FC) layer, which performs a dimensionality reduction to generate the ultimate photovoltaic power generation prediction.

3. EMPIRICAL RESULTS

3.1. Data

This research employs a comprehensive dataset collected from a selected area within Guangdong Province, China containing detailed atmospheric measurements alongside chronological records of PV system performance, including eight dimensions: PV power, total irradiation intensity, temperature, wind speed, air pressure, humidity, total cloud cover, and precipitation. Because the PV power at night is almost zero, this study

FIGURE 6. Pearson correlation analysis.

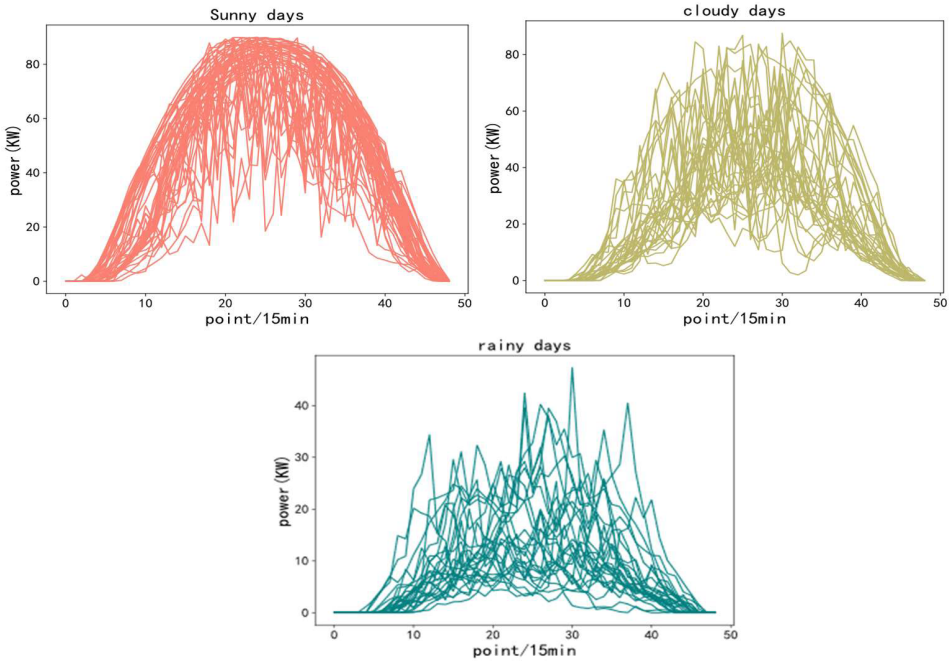

FIGURE 7. Clustering results.

excludes the data at night and processes the anomalous values and missing data points in the data, and finally obtains 5439 valid samples.

The characteristics of meteorological data were screened by Pearson correlation analysis, and the factors strongly correlated with PV power were extracted. The experimental outcomes and corresponding analysis are visually presented in Figure 6.

Total radiant intensity, air temperature, and air pressure exhibit positive correlations with PV power output, whereas wind speed, humidity, and total cloud cover show negative correlations. Specifically, the correlation between total radiant intensity and PV power is the most significant, followed by humidity, and then air temperature. Based on these findings, total radiant intensity, humidity, and air temperature were identified as the primary factors influencing PV power output. These critical variables were consequently selected as fundamental input features for the forecasting framework.

Clustering is based on historical PV data, and Figure 7 provides a visual representation of the clustering results. The analysis of meteorological conditions revealed a distribution of 52 clear-sky days, 27 overcast days, and 32 precipitation days within the study period. On sunny days, the power output of PV generation follows a relatively smooth parabolic trend as solar irradiation intensity changes. In contrast, on cloudy days, increased cloud thickness and variations in total cloud cover enhance atmospheric refraction and reduce solar irradiation, leading to random fluctuations in PV power. On rainy days, high relative humidity and increased water vapor content block ef-

TABLE 1. Main parameters of each model.

Model	Weather	Epoch	Learning rate	Hidden dim
BiLSTM	Sunny Cloudy Rainy	200	1e-3	64
BiLSTM-Bayesian	Sunny	95	9e-4	27
	Cloudy	112	9e-4	95
	Rainy	95	1e-5	27
BiLSTM-Transformer	Sunny Cloudy Rainy	200	1e-3	64
Proposed model	Sunny	95	9e-4	32
	Cloudy	52	1e-3	16
	Rainy	52	9e-4	32

TABLE 2. Evaluating the predictive performance across various models.

Weather	Model	MAE(%)	MSE(%)	RMSE(%)
Sunny	RF	3.62	45.81	6.77
	BiLSTM	3.06	44.42	6.66
	BiLSTM-Bayesian	3.03	28.21	5.31
	BiLSTM-Transformer	0.11	0.03	0.19
	Proposed model	**0.07**	**0.02**	**0.14**
Cloudy	RF	2.34	11.93	3.45
	BiLSTM	3.93	33.53	5.79
	BiLSTM-Bayesian	2.73	14.70	3.83
	BiLSTM-Transformer	0.49	0.68	0.83
	Proposed model	**0.36**	**0.26**	**0.52**
Rainy	RF	5.53	81.59	9.03
	BiLSTM	8.79	182.73	13.52
	BiLSTM-Bayesian	5.51	70.72	8.41
	BiLSTM-Transformer	0.25	0.15	0.38
	Proposed model	**0.22**	**0.14**	**0.37**

fective ground-reflected radiation, resulting in a decrease in PV power output.

3.2. Performance Evaluation

The proposed methodology employs three widely recognized performance metrics for model evaluation: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These quantitative measures are mathematically defined by the following equations, which provide standardized methods for assessing prediction accuracy.

$$MSE = \frac{\sum_{n=0}^{N} (\hat{y}_i - y_i)^2}{N} \tag{16}$$

$$RMSE = \sqrt{\frac{\sum_{n=0}^{N} (y_i - \hat{y}_i)}{N}} \tag{17}$$

$$MAE = \frac{\sum_{n=0}^{N} |y_i - \hat{y}_i|}{N} \tag{18}$$

where y_i represents the actual power measurement at the ith time interval, $\hat{y}_i$ the corresponding predicted value at the same temporal point, and N the complete count of data instances contained within the testing dataset, serving as the normalization factor for the evaluation metric.

3.3. Forecasting results

To evaluate the efficacy of the proposed PV power prediction model, which incorporates the K-means++-BiLSTM-Transformer methodology, comparative experiments are conducted with four other power prediction models, including Random Forest (RF), BiLSTM, BiLSTM-Bayesian, and BiLSTM-Transformer. Default parameters are selected for the RF model. The specific configurations of the models are outlined in Table 1, and Table 2 displays the predictive performance for each respective model.

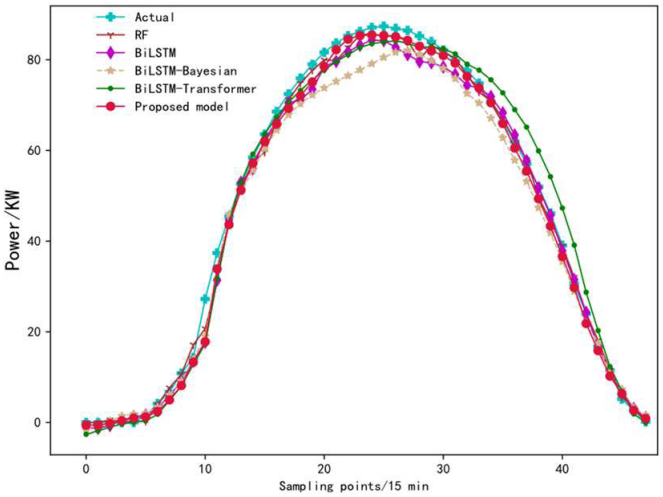

FIGURE 8. Prediction curves under sunny conditions.

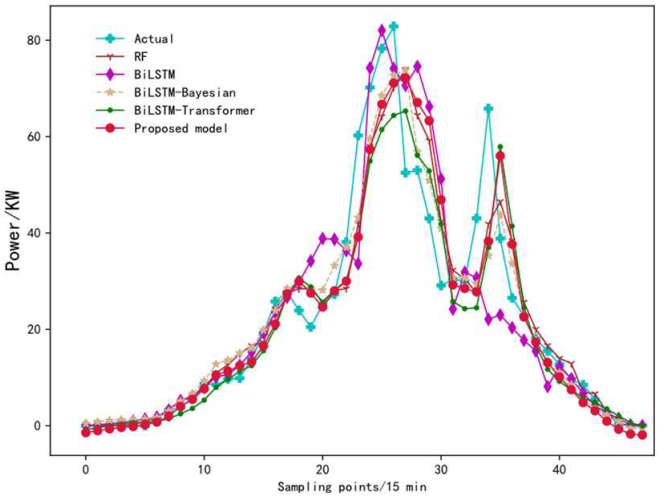

FIGURE 9. Prediction curves under cloudy conditions.

In Table 2, the proposed K-means++-BiLSTM-Transformer combined PV power prediction model outperforms the other three comparison models under three different weather conditions. The K-means++-BiLSTM-Transformer model exhibits low prediction errors and high accuracy. On sunny days, the errors were reduced by 3.55%, 45.79%, and 6.63 compared with RF; 2.99%, 44.40%, and 6.52% compared with BiLSTM; 2.96%, 28.19%, and 5.17% compared with BiLSTM-Bayesian; and 0.04%, 0.01%, and 0.05% compared with BiLSTM-Transformer. Under cloudy conditions, the errors were reduced by 1.98%, 11.67%, and 2.93% compared with RF; 3.57%, 33.27%, and 5.27% compared with the BiLSTM model; 2.37%, 14.44%, and 3.31% compared with the BiLSTM-Bayesian model; and 0.13%, 0.42%, and 0.31% compared with the BiLSTM-Transformer model. On rainy days, the errors are reduced by 5.31, 81.45%, and 8.66% compared with RF; 8.57%, 182.59%, and 13.15% compared with BiLSTM; 5.29%, 70.58%, and 8.04% compared with BiLSTM-Bayesian; 0.03%, 0.01%, and 0.01% less than BiLSTM-Transformer.

To further demonstrate the effectiveness and accuracy of the K-means++-BiLSTM-Transformer combined PV power prediction model under various weather conditions, Figures 7 to 9 present the PV power curves and evaluation index values for the four models, considering diverse weather scenarios. As shown in Figure 8, the data on a sunny day are relatively smooth, with low volatility and stable amplitude, which is more pronounced than rainy days. In terms of overall error data, the K-means++-BiLSTM-Transformer model demonstrates prediction outcomes that are more aligned with the actual values, exhibiting lower error indices than the three other benchmark models.

On a cloudy day, as depicted in Figure 9, the PV power exhibits significant fluctuations with an amplitude slightly lower than that observed on a sunny day. Both the stability and trend are reduced, leading to increased prediction errors. Despite these challenges, the K-means++-BiLSTM-Transformer model maintains its ability to precisely forecast PV power at a majority of sampling points, demonstrating its superior prediction performance.

In the case of rainy weather, as shown in Figure 10, the K-means++-BiLSTM-Transformer model proposed in this study demonstrates the ability to capture the overall trend of changes even under heightened weather complexity, outperforming other models in terms of prediction accuracy. Compared to sunny and cloudy days, the data fluctuates more, and the amplitude is reduced. The results show that other prediction models fluctuate greatly at some sampling points and have a low degree of fitting with active power, indicating that the prediction models cannot capture and process the effective information of the time series well.

Compared with the prediction models BiLSTM and BiLSTM-Transformer without hyperparameter optimization, the optimization of hyperparameters notably enhances the model's performance. Table 2 and Figure 11 demonstrate that the prediction model with Bayesian hyperparameter optimization exhibits lower errors than the BiLSTM model across three distinct weather conditions. For example, in the cloudy case, the indicators of the Bayesian-optimized BiLSTM model decreased by 1.20%, 18.83%, and 5.11%, respectively.

In comparison with the integrated BiLSTM-Transformer forecasting approach, both the conventional RF and individual BiLSTM models show considerably greater inaccuracies in their predictions. As shown in Table 2 and Figure 11, , the hybrid BiLSTM-Transformer framework demonstrates significantly better performance across three distinct meteorological scenarios, achieving substantially smaller prediction errors than individual baseline models. For instance, during clear weather, the integrated system decreases inaccuracies by 2.95%, 44.39%, and 6.47%, respectively, when being measured against the stand-alone BiLSTM framework.

During training on a dataset comprising 2030 data samples, the proposed model takes 113 seconds to complete when using a CPU. In contrast, the stand-alone BiLSTM model requires only 45 seconds. RF requires approximately 13 seconds. Despite the extended training duration of the proposed model relative

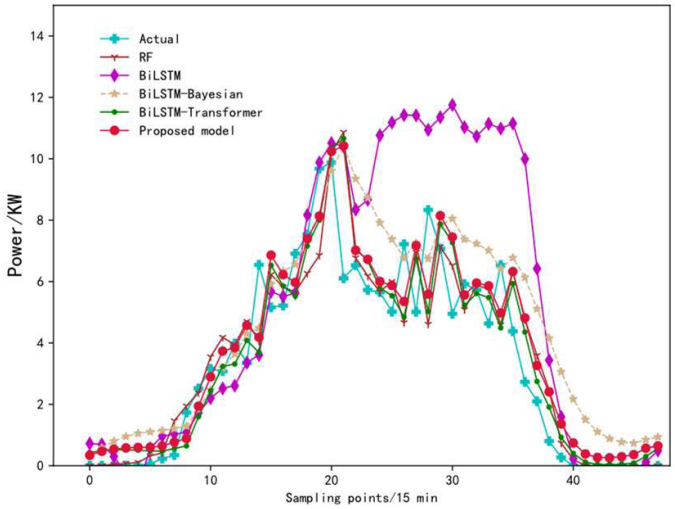

FIGURE 10. Prediction curves under rainy conditions.

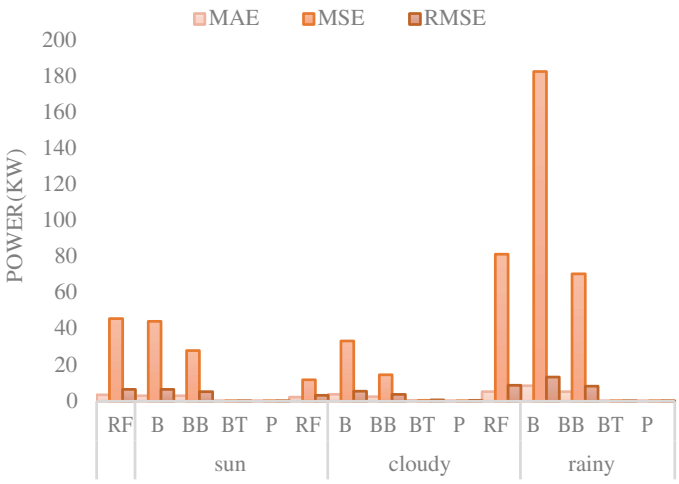

FIGURE 11. Error values under three weather.

to comparative models, empirical results indicate that it attains superior predictive accuracy.

In summary, the K-means++-BiLSTM-Transformer model demonstrates superior performance under both sunny and rainy conditions. It achieves lower MAE, MSE, and RMSE values, thereby validating the high accuracy of the proposed prediction method.

4. CONCLUSIONS

PV power generation is characterized by inherent volatility and uncertainty, posing significant challenges for accurate forecasting. To address these challenges, this study proposes a novel prediction model based on the K-means++-BiLSTM-Transformer framework. The primary objective is to improve the accuracy and reliability of PV power output predictions under diverse weather conditions, leveraging advanced techniques to handle the nonlinear and time-dependent nature of PV data.

(1) The study begins by employing Pearson correlation analysis to identify the three most influential meteorological factors affecting PV output power. This step reduces input data dimensionality and minimizes redundancy. Subsequently, the K-means++ algorithm clusters the data into three distinct weather conditions, enabling the model to adapt to varying environmental scenarios.

(2) The BiLSTM-Transformer model is developed to capture temporal dependencies and complex patterns in the data. The clustered weather data are fed into the model, and Bayesian optimization is applied to fine-tune hyperparameters. This approach ensures parameter optimization tailored to specific meteorological conditions, enhancing model robustness across diverse weather scenarios.

(3) The model's performance is evaluated under three distinct weather conditions using a triad of evaluation criteria. Comparative analysis reveals that the proposed model outperforms existing approaches, particularly in accuracy and reliability for extended temporal sequences. The results demonstrate the model's ability to deliver precise and stable predictions under varying environmental conditions.

The proposed method has been validated solely within a single photovoltaic power station. To enhance the model's generalizability, future research will involve its application and evaluation across diverse photovoltaic power stations.

ACKNOWLEDGEMENT

This paper is supported by the Guangdong Basic and Applied Basic Research Foundation (2022A1515110650) and by Science and Technology Projects in Guangzhou (2024A04J4760).

APPENDIX A

Correlation coefficient	Degree of correlation
0.8–1.0	Very Strong
0.6–0.8	Strong
0.4–0.6	Moderate
0.2–0.4	Weak
0.0–0.2	Very weak

APPENDIX B

Full form	Abbreviations
Random Forest	RF
BiLSTM	B
BiLSTM-Bayesian	BB
BiLSTM-Transformer	BT
Proposed model	P

REFERENCES

[1] Qi, X. and J. Cheng, "A preliminary study on the low carbon development path of China's industry under the double carbon target," *Modern Industrial Economy and Informationization*, Vol. 12, No. 12, 10–12, 2022.

[2] Li, J., Y. Luo, S. Yang, S. Wei, and Q. Huang, "Review of uncertainty forecasting methods for renewable energy power," *High Voltage Engineering*, Vol. 47, No. 4, 1144–1155, 2021.

[3] Hong, T., P. Pinson, Y. Wang, R. Weron, D. Yang, and H. Zareipour, "Energy forecasting: A review and outlook," *IEEE Open Access Journal of Power and Energy*, Vol. 7, 376–388, 2020.

[4] Meng, X., X. Shi, W. Wang, Y. Zhang, and F. Gao, "An upscaling minute-level regional photovoltaic power forecasting scheme," *International Journal of Electrical Power & Energy Systems*, Vol. 155, 109609, 2024.

[5] Yu, J., X. Li, L. Yang, L. Li, Z. Huang, K. Shen, X. Yang, X. Yang, Z. Xu, D. Zhang, and S. Du, "Deep learning models for PV power forecasting: Review," *Energies*, Vol. 17, No. 16, 3973, 2024.

[6] Tina, G. M., C. Ventura, S. Ferlito, and S. D. Vito, "A state-of-art-review on machine-learning based methods for PV," *Applied Sciences*, Vol. 11, No. 16, 7550, 2021.

[7] Barbieri, F., S. Rajakaruna, and A. Ghosh, "Very short-term photovoltaic power forecasting with cloud modeling: A review," *Renewable and Sustainable Energy Reviews*, Vol. 75, 242–263, 2017.

[8] Das, U. K., K. S. Tey, M. Seyedmahmoudian, S. Mekhilef, M. Y. I. Idris, W. V. Deventer, B. Horan, and A. Stojcevski, "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, Vol. 81, 912–928, 2018.

[9] Wang, G., Y. Su, and L. Shu, "One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models," *Renewable Energy*, Vol. 96, 469–478, 2016.

[10] Yang, D., P. Jirutitijaroen, and W. M. Walsh, "Hourly solar irradiance time series forecasting using cloud cover index," *Solar Energy*, Vol. 86, No. 12, 3531–3543, 2012.

[11] Li, Y., Y. He, Y. Su, and L. Shu, "Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines," *Applied Energy*, Vol. 180, 392–401, 2016.

[12] Lee, D. and K. Kim, "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, Vol. 12, No. 2, 215, 2019.

[13] Qin, Y., Y. Xu, X. Wang, T. Wang, and W. Li, "Study on short-term photovoltaic output prediction based on improved FCM-LSTM," *Acta Energiae Solaris Sinica*, Vol. 45, No. 08, 304–313, 2024.

[14] Liu, F. and C. Liang, "Short-term power load forecasting based on AC-BiLSTM model," *Energy Reports*, Vol. 11, 1570–1579, 2024.

[15] Piantadosi, G., S. Dutto, A. Galli, S. D. Vito, C. Sansone, and G. D. Francia, "Photovoltaic power forecasting: A Transformer based framework," *Energy and AI*, Vol. 18, 100444, 2024.

[16] Lin, P., Z. Peng, Y. Lai, S. Cheng, Z. Chen, and L. Wu, "Short-term power prediction for photovoltaic power plants using a hybrid improved Kmeans-GRA-Elman model based on multivariate meteorological factors and historical power datasets," *Energy Conversion and Management*, Vol. 177, 704–717, 2018.

[17] Xu, R. and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, Vol. 16, No. 3, 645–678, 2005.

[18] Jain, A. K., "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, Vol. 31, No. 8, 651–666, 2010.

[19] Celebi, M. E., H. A. Kingravi, and P. A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm," *Expert Systems with Applications*, Vol. 40, No. 1, 200–210, 2013.

[20] Arthur, D. and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *ACM-SIAM Symposium on Discrete Algorithms*, Vol. 7, 1027–1035, 2007.

[21] Balderas, D., P. Ponce, and A. Molina, "Convolutional long short term memory deep neural networks for image sequence prediction," *Expert Systems with Applications*, Vol. 122, 152–162, 2019.

[22] Chen, J., H. Jing, Y. Chang, and Q. Liu, "Gated recurrent unit based recurrent neural network for remaining useful life prediction of nonlinear deterioration process," *Reliability Engineering & System Safety*, Vol. 185, 372–382, 2019.

[23] Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems*, Vol. 30, 2017.

[24] Wang, X., Y. Jin, S. Schmitt, and M. Olhofer, "Recent advances in Bayesian optimization," *ACM Computing Surveys*, Vol. 55, No. 13s, 1–36, 2023.