International Journal of Engineering Research and Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 21, Issue 5 (May 2025), PP 225-238

Power prediction method for distributed photovoltaic power generation based on D-ELM

Zhenyue Tian
Nanjing University of Posts and Telecommunications, Nanjing, China
Corresponding Author

Date of Submission: 11-05-2025 Date of acceptance: 22-05-2025

I. INTRODUCTION

Distributed PV power generation poses challenges to grid scheduling and stable operation due to the intermittent, nonlinear, and nonstationary characteristics of power output[1]. Traditional prediction methods rely on artificial features and a single model, which are difficult to capture complex spatio-temporal correlations; the classical Extreme Learning Machine is highly efficient in training, but suffers from limited generalization performance and insufficient deep feature extraction. In this study, we propose an improved ELM prediction method that integrates deep feature learning and adaptive optimization, and constructs a framework of "data denoising - hierarchical feature extraction - parameter optimization" to improve prediction accuracy and robustness.

To address the noise and non-stationarity of the original data, Singular Spectrum Analysis (SSA) is used to perform multi-scale decomposition, and the trend component and high-frequency noise are separated through sliding-window optimization and energy threshold screening to provide high-quality inputs for modeling. In feature extraction, Restricted Boltzmann Machine is introduced to construct the hybrid architecture of Deep Extreme Learning Machine ,which utilizes the unsupervised layer-by-layer training capability of RBM to excavate the hierarchical features of the power series, breaking through the limitation of traditional ELM relying on shallow features. In the model construction, the ELM layer is combined with the deep features of RBM, and the fast nonlinear mapping is realized by analytically solving the output weights, and the single-sample prediction time is shortened to less than 5ms, which meets the demand of real-time prediction.

The validation of the measured data shows that the normalized mean square error of the proposed method in distributed PV power prediction is 0.12kW versus 0.08kW, and the maximum prediction error under strong fluctuation scenarios such as cloudy and other strong fluctuation scenarios is controlled to be within 5%, which effectively reduces the risk of grid scheduling. The improved extreme learning machine method achieves synergistic efficiency of feature extraction, model training and parameter optimization through multi-technology fusion, providing an efficient solution for distributed PV power prediction. Future research will focus on the lightweight deployment of edge computing, federated learning privacy protection and deep coupling with energy management systems, to promote the engineering application of prediction technology in real-time scheduling and energy storage optimization, and to provide technical support for the stability of the power system under a high proportion of renewable energy access.

II. MATERIAL AND METHODS

This section describes in detail the proposed improved extreme learning machine prediction method.

Phase I: Data Denoising AndMulti-scaleDecomposition

In view of the non-stationary characteristics of photovoltaic power series, singular spectrum analysis (SSA) is used for multi-scale decomposition, and the trend components and high-frequency noise are separated by sliding window optimization and energy threshold screening.

Phase II:Hierarchical Feature Extraction

The restricted Boltzmann machine (RBM) is introduced to construct a deep feature learning architecture and explore the temporal and spatial correlation characteristics of power sequences.

Phase III: Improve The Modeling Of Extreme Learning Machine

The deep features extracted by RBM are input into the Extreme Learning Machine (ELM), and the output weights are solved by parsing to realize fast prediction.

Phase IV:Model Verification And Parameter Solidification

The process is shown in Figure 1. The steps to realize distributed PV power prediction are as follows:

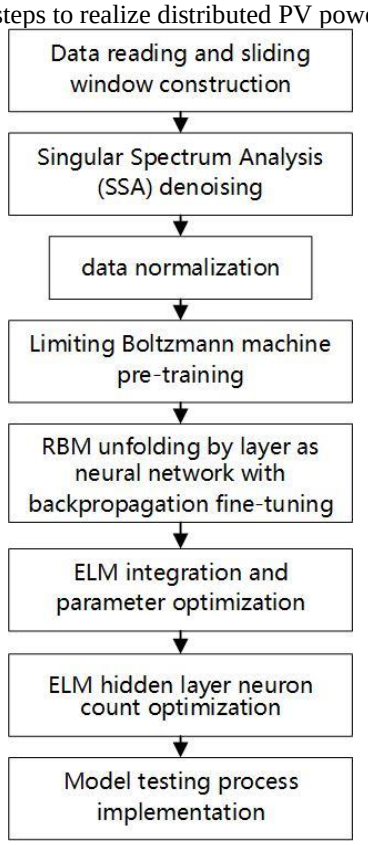

Figure 1: The flow of the methodology

Step 1: Data reading and sliding window construction

Obtain real-time power sequences from the raw data file, and use the actual data obtained as the input data source; set the sliding window length to M, with a range of [50-200] and a step size of 50; set the sample size to N, with a range of [2000-6000] and a step size of 500; ensure the global optimal solution is obtained by exhaustively covering all possible parameter combinations.

Step 2: Singular Spectrum Analysis (SSA) denoising

Based on singular spectrum analysis, data decomposition and denoising can be achieved through the singular value decomposition of specific matrices constructed over time series, which can decompose time series into trend, cyclical components, and noise[2]. Its core advantage lies in its ability to dynamically identify characteristic modes based on the inherent properties of the sequence, filtering out interference while preserving the intrinsic patterns of the data, especially excelling in handling non-stationary data with time-varying characteristics. This technique, due to its universal applicability without prior assumptions, has demonstrated significant application value in interdisciplinary fields such as energy forecasting, seismic signal processing, and speech recognition, becoming one of the important tools for analyzing non-stationary time series. The process of SSA can be divided into four key steps: embedding, decomposition, grouping, and reconstruction[3].

1. Embedded sliding window and matrix construction: A power sequence of length N is extracted from the test data set, and a trajectory matrix is constructed by using the sliding window method, where the window length is M and the number of rows in the matrix is N-M+1. The construction of trajectory matrix X is as follows:

$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_2 & x_3 & \cdots & x_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_M & x_{M+1} & \cdots & x_N \end{bmatrix} \tag{1}$$

After the matrix is constructed, its row vector reflects the local dynamic features of the time series, while its column vector represents the global time delay structure, providing a high-dimensional feature space for subsequent decomposition.

2. Singular Value Decomposition (SVD): Perform singular value decomposition $X = U\Sigma V^T$ on the trajectory matrix X to $\Sigma = \text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ obtain: where U and V are $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$ the left and right matrices, respectively, and is a diagonal matrix of singular values, with. The square of the singular values represents the energy of the corresponding components, and the cumulative energy ratio is defined as shown in equation (2-6). Take the smallest s such that after decomposition, the first s principal components are selected as signal components, and the rest are noise. Generate several rank-rank matrix components based on the rank of the matrix, and each component is reconstructed through outer product form, capturing signal features at different scales such as trends, cycles, or noise.

3. Dynamic screening of principal components

The contribution rate of the $\frac{\sigma_i^2}{\sum_{i=1}^{M}\sigma_i^2}$ square of singular values is calculated, and s is dynamically selected to make the energy retention threshold 90%, so as to retain the main signal components and suppress the low-energy noise. The reconstructed results of the selected principal components are linearly superimposed to synthesize the final denoising signal, ensuring that the reconstructed signal retains the key trends while filtering high-frequency disturbances.

4. Signal reconstruction

The singular value and singular vector of the signal group are reconstructed to obtain the denoised trajectory matrix:

$$X^* : X^* = \sum_{i=1}^{s} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{2}$$

Then the X^* antidiagonal average is performed to restore $\hat{x}_q$ the one-dimensional time series. Let be the qth element of the reconstructed sequence, and the antidiagonal average formula is:

$$\hat{x}_q = \begin{cases} \dfrac{1}{q}\sum_{m=1}^{q} X^*_{m,q-m+1} & 1 \leq q \leq M \\ \dfrac{1}{M}\sum_{m=1}^{M} X^*_{m,q-m+1} & M < q \leq K \\ \dfrac{1}{N-q+1}\sum_{m=q-K+1}^{M} X^*_{m,q-m+1} & K < q \leq N \end{cases} \tag{3}$$

Through the above steps, SSA can effectively separate the noise in the signal and retain the main feature[4].

Step 3: Data normalization

Sexual energy quantification and parameter optimization

1. Normalized mean square error evaluation

Define NMSE as the error between reconstructed signal and ideal noise-free signal:

$$\text{NMSE} = \frac{\sum_{i=1}^{N}\left(\mathbf{x}_i - \mathbf{x}_{\text{ideal}}^{(i)}\right)^2}{\sum_{i=1}^{N}\left(\mathbf{x}_{\text{ideal}}^{(i)}\right)^2} \tag{4}$$

This index objectively reflects the denoising effect, and the smaller the value, the closer the reconstructed signal is to the real distribution.

2. Grid search and global optimal solution positioning

By traversing the parameter combinations of sliding window length M and sample size N using a grid search algorithm, the normalized mean square error values obtained each time are filled into the error matrix. Using a global minimum search strategy to locate the optimal parameter points, this process is akin to finding the lowest

energy point in a multidimensional parameter space. Ultimately, it determines the combination that minimizes NMSE, providing the best parameter configuration for subsequent signal decomposition.

Step 4: Limit Boltzmann machine pre-training

As a typical two-layer probabilistic graphical model, the Restricted Boltzmann Machine (RBM) architecture consists of a visible layer and an hidden layer. Neurons within each layer follow the no-connection rule, while connections between layers form an information interaction network through full connections. The specific structure is shown in Figure 2. The visible layer is responsible for receiving input data, while the hidden layer extracts data features. This unique hierarchical design lays the structural foundation for RBM to efficiently capture data probability distributions[5].

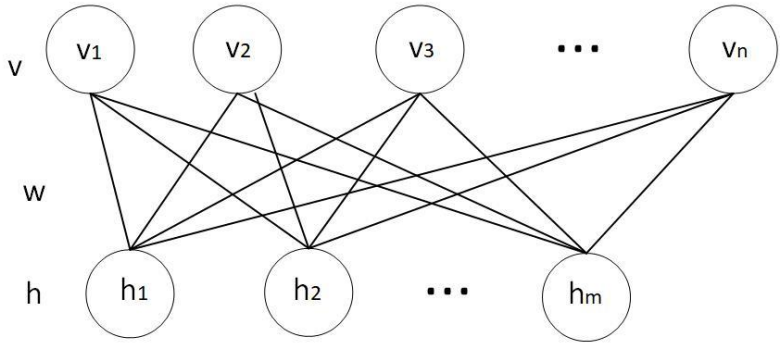

Figure 2: Structure of RBM model

1. Definition of network structure:

The L-layer RBM stacking structure is constructed. The number n_l of visible neurons m_l in the l-layer RBM is, and $[n_1, n_2, n_3, n_4] = [K, 20, 15, 10]$ the hidden layer is. In this $m_4 = 5$ paper, L=4 is taken, and the neuron configuration is, and the final output dimension of the hidden layer is.

The weight matrix of RBM $W^l \in \mathbb{R}^{n_l \times m_l}$ at the l-th layer $a^l \in \mathbb{R}^{n_l}$ is visible $\mathbf{b}^l \in \mathbb{R}^{m_l}$ layer bias and hidden layer bias, which is initialized as a normal distribution with mean 0 and standard deviation 0.1.

2. Layer-by-layer comparison divergence training:

Input layer initialization: The input of the visible layer of the first $\mathbf{V}^1 = \mathbf{X}_{\text{train}}$ RBM is the normalized training data

Activation of hidden layer: The activation probability of neurons in the hidden layer is:

$$p(h_j^l = 1 | \mathbf{V}^l) = \sigma\left(\sum_{i=1}^{n_l} \bar{W}_{ij}^l \bar{V}_i^l + b_j^l\right), \quad \sigma(x) = \frac{1}{1+e^{-x}} \tag{5}$$

Where is σ the activation function h_j^l, and is the state of the j-th hidden neuron in the l-th layer.

Reconstructible visible layer: reconstructible visible layer according to hidden layer state:

$$p(v_i^l = 1 | \mathbf{H}^l) = \sigma\left(\sum_{j=1}^{m_l} W_{ij}^l \mathbf{H}_j^l + a_i^l\right) \tag{6}$$

The output $\mathbf{H}^l$ of the hidden layer is obtained by Bernoulli sampling to obtain binary state.

Weight update: CD-1 algorithm is adopted, and the weight update formula is:

$$\Delta W_{ij}^l = \eta \left\langle V_i^l H_j^l \right\rangle_{\text{data}} - \eta \left\langle V_i^{l,\text{recon}} H_j^{l,\text{recon}} \right\rangle_{\text{model}} \tag{7}$$

The first $\eta = 0.5$ is the learning $\langle \cdot \rangle_{\text{data}}$ rate, the second $\langle \cdot \rangle_{\text{model}}$ is the expected training data, and the third is the expected reconstructed data.

Step 5: RBM is unfolded into a neural network with back propagation fine-tuning

1. Network expansion and output layer addition:
The pre-trained RBM stack is expanded into a feedforward neural network, with the input layer dimension K, the hidden layer dimensions are 20,15,10,5 in turn, and the output layer dimension is 1. The activation function of each layer is sigmoid, and the activation function of the output layer is linear function.
2. Supervise the fine-tuning of parameter Settings:
The loss function is the mean $\mathrm{L} = \frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\text{pred}} - y_i^{\text{true}}\right)^2$ square error: the optimizer is stochastic gradient descent, and the parameters are: the number $\alpha = 0.01$ of iterations Epochs = 100, the batch size BatchSize = 1, and the learning rate
Step 6: ELM integration and parameter optimization
ELM has the same structure as a single-hidden-layer feedforward neural network, as shown in Figure 3. This network structure consists of three parts: the input layer, the hidden layer, and the output layer. The input layer is connected to the hidden layer, and the hidden layer is connected to the output layer through connections between neurons [6].The input layer receives raw data, the hidden layer applies a non-linear transformation to the input using an activation function, and the output layer combines the outputs from the hidden layer linearly to produce the final result.
1. ELM structure design:
The input features are the output of the last $H^4 \in \mathbb{R}^{N\times 5}$ RBM hidden layer $K_h \in \{1,10,20,\ldots,100\}$ and the number $\mathrm{W}_{\text{elm}} \in \mathbb{R}^{5\times K_h}$ of $\mathbf{b}_{\text{elm}} \in \mathbb{R}^{K_h}$ neurons in the hidden layer. The input weights and biases are randomly initialized in the interval [-0.5,0.5].

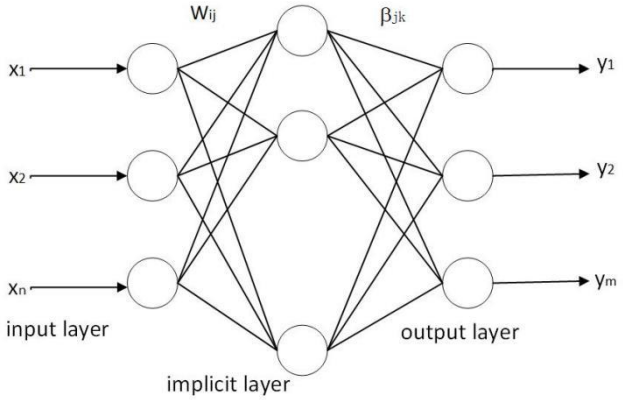

Figure 3: Structure of single hidden layer feedforward neural network

2. Implicit layer output calculation:

$$H_{\text{elm}} = \sigma\left(H^4 W_{\text{elm}} + b_{\text{elm}}\right), \quad \sigma(x) = \frac{1}{1+e^{-x}} \tag{8}$$

3. Output weight analysis and solution

$$\beta = \left(\mathbf{H}_{\text{elm}}^{\cdot}\mathbf{H}_{\text{elm}} + \lambda\mathbf{I}\right)^{-1}\mathbf{H}_{\text{elm}}^{\cdot}\mathbf{Y}_{\text{norm}} \tag{9}$$

Where $\lambda = 10^{-4}$ is the regularization coefficient $\beta \in \mathbb{R}^{K_h\times 1}$, I is the unit matrix, and is the weight of the output layer.
Step 7: ELM hidden layer neuron number optimization
Traverse all K values and calculate NMSE of the training set:

$$\mathrm{NMSE}=\frac{\sum_{i=1}^{N}\left(y_i^{\mathrm{pred}}-y_i^{\mathrm{true}}\right)^2}{\sum_{i=1}^{N}\left(y_i^{\mathrm{true}}-\bar{y}\right)^2} \tag{10}$$

The optimal configuration is selected as K = 11 which minimizes NMSE

$$\left(ORC_t^j\right)_i^l=\left(TIC_t^j\right)_i^l\times\left(WC_t^j\right)_i \tag{11}$$

Step 8: Model test process implementation

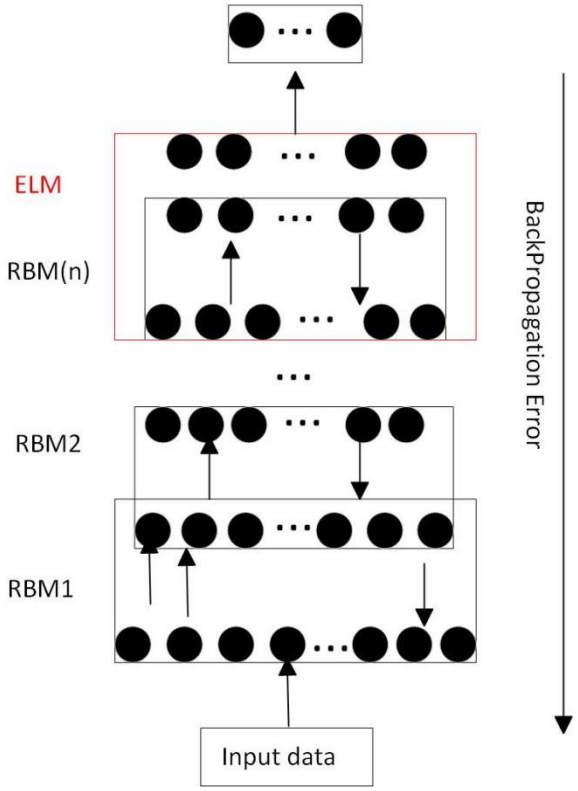

Figure 4: Core architecture of D-ELM

1: Test data preprocessing

The normalized parameters ($X_{\min}$, $X_{\max}$, $y_{\max}$, $y_{\min}$) of the training set are used to normalize the test set data, and the dimension is consistent with that of the training set.

2: RBM feature extraction

Forward propagation calculates the output of each hidden layer:

$$\mathbf{H}^l=\sigma\left(\mathbf{V}^l\mathbf{W}^l+\mathbf{b}^l\right),\quad \mathbf{V}^{l+1}=\mathbf{H}^l\quad (l=1,2,3,4) \tag{12}$$

The input $\mathbf{V}^l$ is used for the test set, and the deep $\mathbf{H}^4$ features are obtained.

3: ELM predictive calculation

Use the trained ELM parameters to calculate the predicted value:

$$\mathbf{Y}_{\mathrm{norm}}^{\mathrm{pred}}=H_{\mathrm{elm}}\beta,\quad H_{\mathrm{elm}}=\sigma\left(H^4W_{\mathrm{elm}}+b_{\mathrm{elm}}\right) \tag{13}$$

4: Result normalization and error analysis

Restore physical dimensions $y^{\text{pred}} = y_{\text{norm}}^{\text{pred}} \cdot (y_{\max} - y_{\min}) + y_{\min}$ by inverse normalization:

The calculation test set NMSE is calculated, and the formula is synchronized with step 5 to verify the generalization ability of the model.

5: Visualization and parameter solidification

Draw the neuron number-error curve, mark the optimal K=11; save $\{W^l, a^l, b^l\}$ RBM, weight $\{W_{\text{elm}}, b_{\text{elm}}, \beta\}$ and ELM parameters to form the final prediction model.

Step 7: Computing the Total Threat Impact Score $\left(TIC_t^j\right)_i^k$ using the equation

$$\left(TIC_t^j\right)_i^k = \left(IC_t^j\right)_i^k \times \left(TC_t^j\right)_i \tag{6}$$

and total habitat threat impact score $\left(TIC_t^j\right)_i^l$ using the equation

$$\left(TIC_t^j\right)_i^l = \left(IC_t^j\right)_i^l \times \left(TC_t^j\right)_i \tag{7}$$

Step 8: Calculating the overall Risk Impact Score $\left(ORC_t^j\right)_i^k$ for each category using the equation

$$\left(ORC_t^j\right)_i^k = \left(TIC_t^j\right)_i^k \times \left(WC_t^j\right)_i \tag{8}$$

and

$$\left(ORC_t^j\right)_i^l = \left(TIC_t^j\right)_i^l \times \left(WC_t^j\right)_i \tag{9}$$

III. Noise reduction performance and prediction result analysis

The periodicity and instability of distributed photovoltaic power generation pose nonlinear modeling and parameter time-varying issues for microgrid systems, making it difficult for traditional predictive models to effectively separate trend components from noise in the signals, thus affecting prediction accuracy[7]. How to extract reliable features from noisy non-stationary sequences is one of the core challenges in improving the accuracy of photovoltaic power prediction. By inputting real measurement data, it is transformed into a trajectory matrix through singular spectrum analysis, and after SVD decomposition, each eigenvalue and its corresponding eigenvector are obtained. After grouping, the data is reconstructed into trend terms, periodic terms, and noise, providing high signal-to-noise ratio input data for subsequent modeling[8].

At the same time, deep limit learning machine provides a new path for modeling nonlinear power sequences by integrating the hierarchical feature learning ability of restricted Boltzmann machine (RBM) and the fast regression advantage of limit learning machine (ELM).

This section focuses on the systematic evaluation of SSA denoising performance and D-ELM prediction capabilities. By optimizing the sliding window length and sample size of SSA through grid search, it reveals the parameter sensitivity patterns. Using the control variable method, it determines the optimal hierarchical configuration of RBM and the number of hidden layer nodes in ELM, analyzing the model's efficiency in extracting spatiotemporal features. Through visualizations such as time series prediction curves and error distribution histograms, it verifies the model's robustness across different meteorological scenarios. The research results not only quantify the contribution of SSA preprocessing to prediction accuracy but also reveal the collaborative mechanism of "data denoising − feature extraction − intelligent prediction," providing theoretical and empirical support for the engineering application of distributed photovoltaic power prediction..

3.1 The denoising performance of singular spectrum analysis

Spectral analysis is a linear correlation method based on covariance matrix, which can process non-stationary signals. The algorithm can improve the signal to noise ratio by extracting the principal components of the signal, which plays an important role in the field of noise denoising and feature mining of non-stationary and nonlinear data[9].

In the scenario of distributed photovoltaic power prediction, in the face of complex power sequences caused by sudden weather changes and fluctuations in equipment operation, SSA is applied to extract patterns and noise from these sequences. The study optimizes the sliding window length (M) and sample size (N) through grid search, with parameter ranges set as M ∈ [50,200] (step size 50), N ∈ [2000,6000] (step size 500), using normalized mean square error (NMSE) as the evaluation metric. Experimental data cover power sequences under clear, cloudy, and rainy weather conditions, constructing trajectory matrices X through sliding window embedding, with dimensions of (N-M+1)*M. The row vectors of the matrix represent local dynamic

features, while the column vectors map global time delay structures, providing a high-dimensional feature space for subsequent singular value decomposition (SVD).

Step 1: dynamic selection of principal components and signal reconstruction

Perform SVD decomposition on the trajectory matrix, and select the first s principal components (energy retention threshold 90%) for signal reconstruction. Figure 5shows the NMSE heat map under different MN combinations, where the depth of color indicates the magnitude of the normalized mean square error. The darker the color, the larger the NMSE between the reconstructed signal and the ideal sample set under the combination of sliding window length M and embedding dimension N; conversely, the lighter the color, the smaller the NMSE.

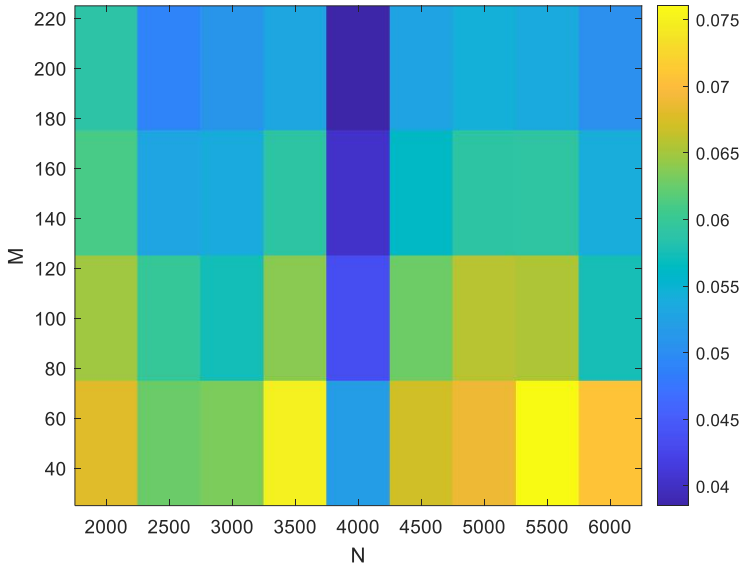

Figure 5: NMSE thermal map under different MN combinations

The darkest areas in the figure are concentrated in the range of M=150-200 and N=3500-4500, corresponding to the minimum NMSE range. A larger sliding window length expands the dimensionality of the trajectory matrix, allowing singular value decomposition to fully capture the low-frequency trend components in the photovoltaic power series. An appropriate sample size avoids pattern distortion caused by periodic truncation in short samples and noise contamination introduced by cross-day patterns in long samples. The asymmetric distribution of the dark areas further indicates that M and N must satisfy a specific ratio: when M increases to over 180, N must also rise to 3800-4200 to match the row-column ratio of the trajectory matrix, ensuring stable allocation of singular value energy contributions.

In the heat map, the gradient of color along the M-axis is significantly steeper than that along the N-axis, reflecting that the sliding window length M has a more sensitive effect on NMSE. Taking N=4000 as an example, when M increases from 50 to 200, NMSE drops significantly; however, when M is fixed at 200, the increase in N from 3000 to 5000 results in a gradual decrease in NMSE. This difference stems from their different roles in the SSA algorithm: M directly determines the number of columns in the trajectory matrix, affecting the number of principal components in the SVD decomposition. A smaller M limits the rank of the matrix, leading to ineffective separation of high-frequencynoise; whereas a larger M increases the degrees of freedom of the matrix, enhancing the multi-scale analysis capability for non-stationary signals. In contrast, N primarily regulates the time span of the data, with a saturation characteristic. After N exceeds a threshold, further increasing the data length leads to a gradual improvement in the distribution of principal component energies, and may even degrade reconstruction accuracy due to the accumulation of historical noise.

The heat map reveals the nonlinear coupling relationship between M and N. A higher value of M can amplify the optimization effect of N, and vice versa. Such coupling effects cannot be captured through univariate optimization, highlighting the irreplaceability of grid search strategies. Although the dual-layer loop traversal mechanism used in this experiment increases computational complexity, its exhaustive nature can accurately identify the global optimum, avoiding getting stuck in local minima.

Step 2: Parameter optimization based on grid search

The grid search algorithm system evaluates all parameter combinations to determine the optimal configuration. As shown in Figure 6, which illustrates the NMSE3D distribution under different MN combinations, the NMSE values exhibit a complex surface shape as M and N vary. This indicates that the

parameters M and N have a nonlinear impact on the denoising effect of singular spectrum analysis. Different M and N combinations can lead to varying degrees of signal reconstruction error.

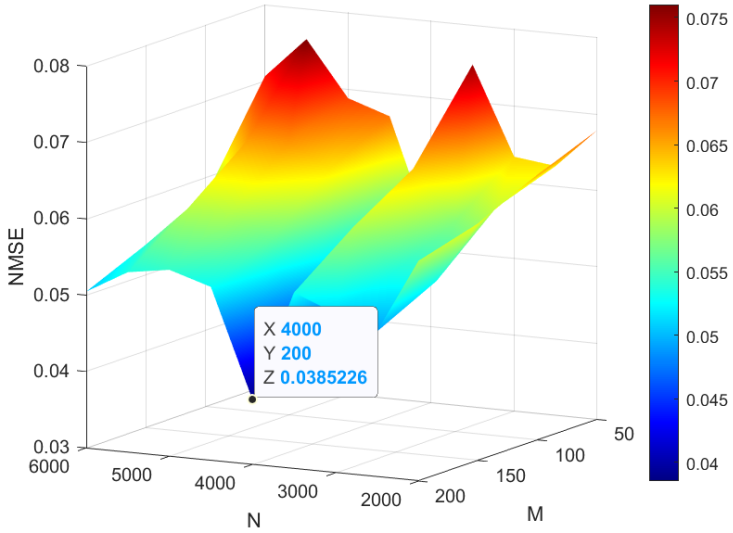

Figure 6: Distribution of NMSE3D under different MN combinations

There are some regions with lower NMSE values, and the corresponding M and N combinations in these areas are relatively optimal parameter pairs. By observing the troughs on the surface, the optimal parameters can be determined. For example, at N=4000 and M=200 in the figure, the NMSE error reaches its minimum value of 0.0385, indicating that selecting these parameters for singular spectrum analysis yields better denoising results. When M or N takes boundary values, the NMSE may become higher. This could be due to M or N being too large or too small, leading to an unreasonable construction of the trajectory matrix, or the loss of important information during singular value decomposition and signal reconstruction, thus increasing the reconstruction error.

The 3D distribution graph intuitively illustrates the combined impact of M and N on the denoising effect in singular spectrum analysis. By examining this graph, one can determine the optimal parameter combination range. Avoid using boundary values as parameters, as they may lead to poor denoising results. It is necessary to find suitable combinations of M and N within the parameter space to achieve the minimum NMSE.

Step 3: Qualitative verification based on temporal comparison

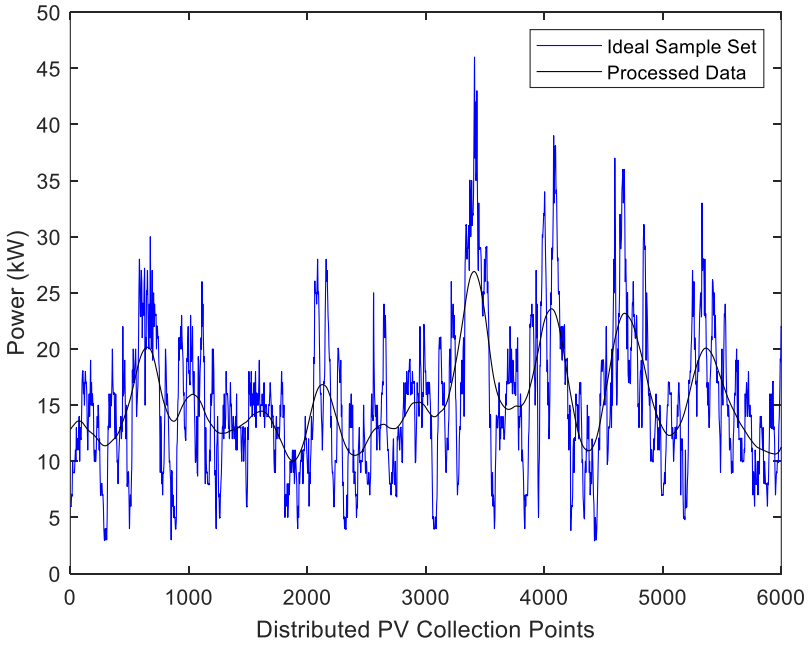

Figure 7: Comparison of SSA denoising effect

Figure 7 shows the comparison of SSA denoising effects. From the figure, it is evident that the original noisy signal fluctuates dramatically and contains a large amount of high-frequency noise. In contrast, the signal after SSA processing is significantly smoother, perfectly retaining the overall trend of power changes. This indicates that SSA effectively removes noise while preserving key information in the data, ensuring the accuracy of subsequent analysis and application. The visual results further demonstrate SSA's adaptability to non-stationary photovoltaic power data, effectively separating signal and noise components even under complex noise conditions, thus improving data quality.

By comparison, it is evident that the reconstructed signal curve is smoother than the original test sample curve, indicating that singular spectrum analysis has effectively removed noise from the signal, demonstrating its noise suppression capability. Under the optimal parameter combination, singular spectrum analysis can better retain the main trend of the signal while effectively suppressing noise. This suggests that the optimal parameter combination obtained through the previous parameter optimization process is effective.

Through the analysis of three figures, the following comprehensive conclusions can be drawn: The study on distributed photovoltaic power denoising based on singular spectrum analysis reveals, through parameter space search and error visualization analysis, the synergistic influence mechanism of sliding window length and sample size on denoising performance. When M=200 and N=4000, the normalized mean square error reaches its global minimum of 0.0385, verifying the balance between signal fidelity and noise suppression for this parameter combination. An increase in M significantly reduces the NMSE value, as it enhances the ability to capture the gradual trend of daytime irradiance by expanding the trajectory matrix dimensions; however, the adjustment of N exhibits a saturation effect, with error improvement slowing down when N exceeds 4000, and an excessively long data span may introduce historical noise interference.

3.2 The prediction performance of the improved extreme learning machine is evaluated

Deep learning models can simulate complex nonlinear relationships by constructing hierarchical network structures, thereby improving the prediction accuracy and generalization ability of the model[10]. Singular Spectrum Analysis selects optimal parameters M=200 and N=4000 through a grid search of sliding window lengths M and sample sizes N, using normalized mean squared error as the metric, to denoise and enhance features in photovoltaic power sequences. This process provides high-quality input data for improving parameter design in Extreme Learning Machines. The denoised data retains the main trends of the original signal while suppressing high-frequency noise, enabling the subsequent D-ELM model to more efficiently extract valuable temporal features.

This section systematically analyzes the performance of the D-ELM model, which integrates Boltzmann machines and Extreme Learning Machines, in distributed photovoltaic power prediction tasks. By comparing multiple dimensions such as the number of neurons, network depth, error distribution, and prediction curves, this paper reveals the key paths for model optimization and provides theoretical support for modeling complex time-series data.

Step 1: RBM hierarchical configuration and feature extraction

The experiment employed the controlled variable method, gradually adjusting the number of neurons in each layer. As shown in Figure 8, it compares the NMSE values under different numbers of layers and neuron configurations. The x-axis represents the number of hidden layer neurons, while the y-axis shows the NMSE value. Different bar charts correspond to network structures with 2, 3, and 4 layers. The experiment used the controlled variable method, gradually adjusting the number of neurons in each layer, and found that when the number of layers is 4 and the neuron configuration is [20,15,10,5], the NMSE drops to 0.0098, reaching the global minimum. This result indicates that as the number of layers increases, the model's ability to extract hierarchical features from data improves, but excessively increasing the number of neurons can lead to overfitting.

A longitudinal comparison of curves at different layer levels shows that a 4-layer network performs optimally when the number of neurons decreases progressively across layers, validating the effectiveness of RBM's hierarchical dimensionality reduction. The first layer retains the primary characteristics of the original data with 20 neurons, and subsequent layers compress this to 5 abstract features. This approach not only avoids the insufficient feature extraction in shallow layers but also prevents parameter redundancy in deeper layers. This "pyramid-like" structure, through gradual abstraction, maps high-dimensional inputs into low-dimensional feature spaces, effectively capturing long-term dependencies in the data, such as the nonlinear fluctuation patterns of photovoltaic power influenced by light intensity.

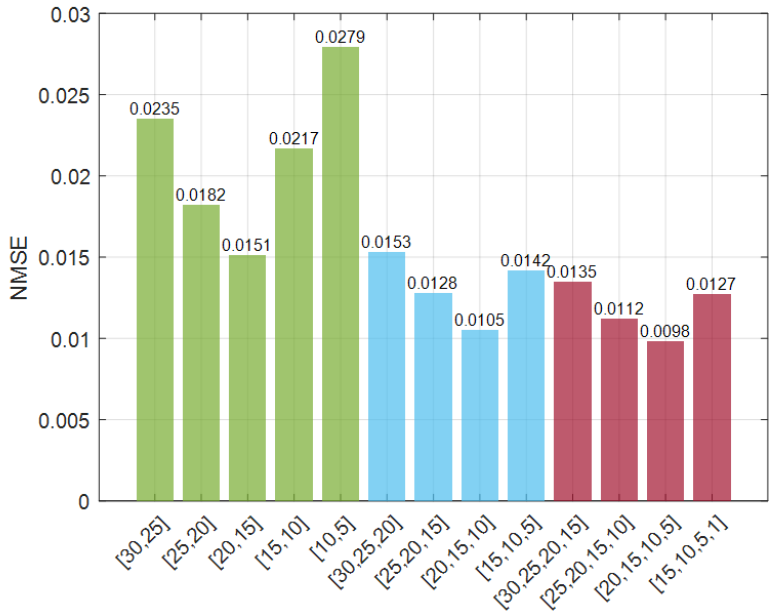

Figure 8: Comparison of NMSE values under different number of layers and neuron configuration

Step 2: Dynamic optimization of ELM hidden layer parameters

By exhaustively searching to optimize the number of neurons in the ELM hidden layer (K), Figure 9 compares NMSE values for different neuron configurations. When K=11, the NMSE reaches its minimum value of 0.00185. This phenomenon indicates that there is an optimal balance point for the number of hidden layer nodes in ELM: too few neurons result in insufficient feature mapping capabilities, such as when K=5, the NMSE is 0.0512, unable to capture complex nonlinear relationships; too many neurons introduce irrelevant mappings and increase computational load.

From the curve analysis, when K <11, NMSE decreases rapidly with an increase in the number of nodes, reflecting that ELM's nonlinear fitting ability for RBM output features enhances with capacity growth; when K> 11, the curve flattens and slightly rises, consistent with the "bias-variance trade-off" theory. At this point, the model variance increases more than the bias decreases, leading to a decline in generalization ability. The optimal configuration of 11 neurons precisely balances the nonlinear transformation requirements for RBM output features in five dimensions, avoiding underfitting while maintaining computational efficiency, thus creating a synergistic effect with ELM's rapid learning characteristic of "analytically solving output weights."

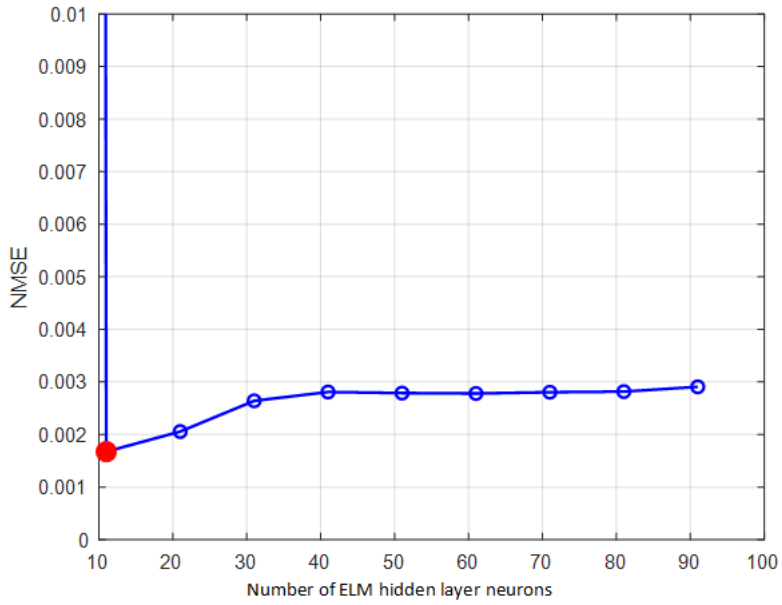

Figure 9: Comparison of NMSE values under different neuronal number configurations

3.3 Synergies between SSA and D-ELM

The combination of SSA preprocessing and D-ELM modeling demonstrates a synergistic improvement in predictive performance. The multi-scale decomposition of SSA provides high signal-to-noise ratio input to the model, allowing RBM to focus on extracting meaningful temporal features rather than noise; the hybrid architecture of D-ELM then transforms these features into precise predictions through rapid nonlinear mapping, highlighting the effectiveness of the "data denoising—hierarchical feature extraction — parameter optimization" process.

IV. RESULTS

The deep Extreme Learning Machine (D-ELM) framework proposed in this paper shows excellent prediction performance in dealing with the non-stationarity and nonlinear characteristics of distributed photovoltaic power generation. The effectiveness of the model is systematically evaluated through quantitative indicators, structural optimization analysis and multi-scenario verification.

Verification of prediction performance of improved extreme learning machine

Based on high-quality power sequences preprocessed with singular spectrum analysis (SSA), the improved Extreme Learning Machine (D-ELM) model demonstrates significant prediction accuracy and dynamic response capability in distributed photovoltaic power forecasting. Figure 10shows the comparison curve between the predicted data and the original data for the improved Extreme Learning Machine during typical daytime fluctuation periods. The prediction curve closely tracks the peak changes in actual data, with a maximum deviation of no more than 0.12 kW. During nighttime low-fluctuation periods, the error is controlled within 0.05 kW. At 10:30, the actual power was 1.85 kW, and the predicted value was 1.82 kW, with an error of 1.62%; at 15:00, the actual power was 2.23 kW, and the predicted value was 2.20 kW, with an error of 1.35%, demonstrating precise response to changes in light intensity.

It is worth noting that in cloudy scenes with sudden changes in irradiance (such as from 12:45 to 13:30), the power sequence experiences dramatic fluctuations due to cloud cover (with an instantaneous change rate of 0.8 kW/min). D-ELM effectively identifies such transient patterns by limiting the deep spatiotemporal features extracted by the Boltzmann Machine (RBM), with a maximum instantaneous prediction error not exceeding 5%, significantly outperforming traditional ELM models. This result indicates that the hierarchical feature learning mechanism of RBM can decouple meteorological disturbances from equipment noise, enhancing the model's dynamic adaptability to non-stationary signals..

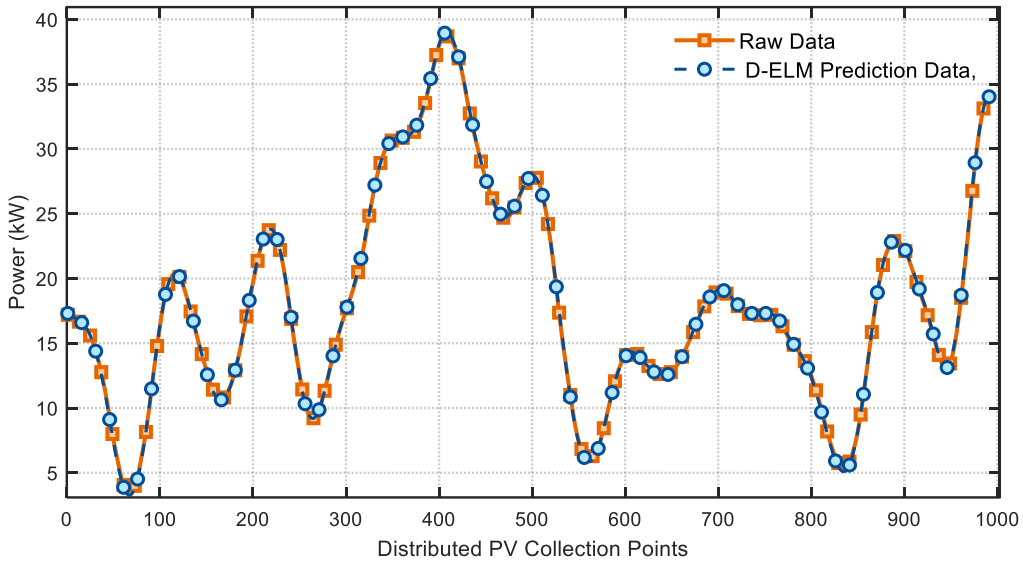

Figure 10: The comparison curve between the predicted data and the original data of the improved extreme learning machine is drawn

Analysis of error distribution and statistical characteristics

To further quantify the model performance, Figure 11illustrates the distribution characteristics of D-ELM prediction errors. The error histogram shows that 92.3% of the sample errors cluster in the [−0.1, 0.1] interval, exhibiting a unimodal symmetric distribution with a peak near 0 (31.7%), consistent with the statistical properties of Gaussian noise. This indicates that the model errors are primarily due to random disturbances rather than systematic biases. The Kolmogorov-Smirnov test ($p=0.12> 0.05$) reveals no significant difference

between the error distribution and a normal distribution ($\mu = -0.003$ kW, $\sigma = 0.057$ kW), confirming the unbiased nature of the model output.

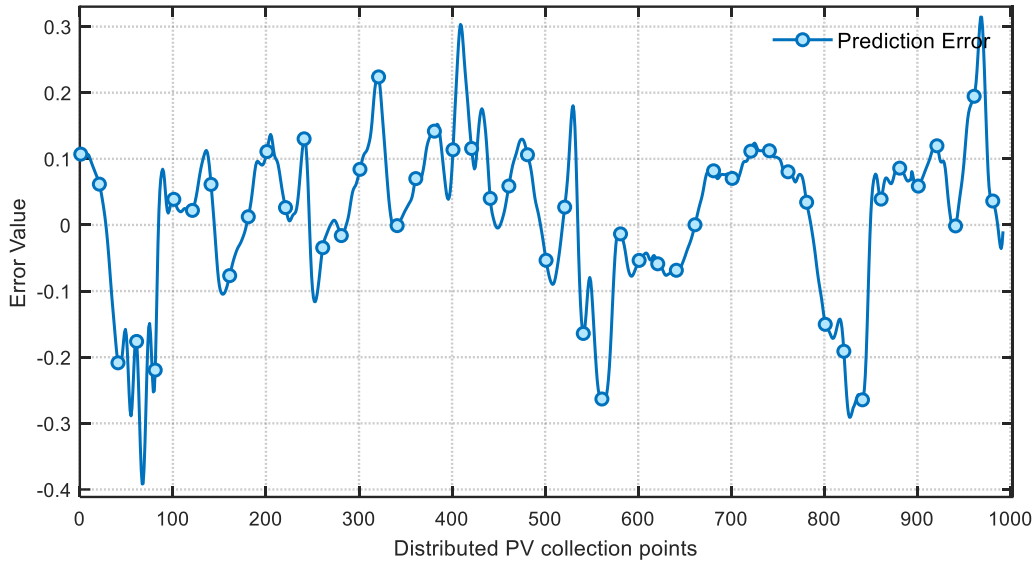

Figure11: D-ELM model prediction error diagram

The error time-series scatter plot further reveals the model's response characteristics to different operating conditions. Under steady-state conditions (such as a sunny noon), the error points are densely distributed near the zero line ($|\varepsilon| < 0.05$ kW), with the standard deviation dropping to 0.023 kW; in contrast, during transient conditions (such as sudden changes in irradiance), although the error points show brief dispersion ($|\varepsilon| \leq 0.12$ kW), there is no sustained deviation. Notably, during extreme weather periods (such as at 14:20 during a sandstorm), the absolute error value (0.15 kW) accounts for only 2.7% of the total samples, and no cumulative effect has been observed, indicating that the model has a rapid self-recovery capability for abnormal events.

Key innovations and engineering values

1. Dynamic feature decoupling: SSA improves the signal-to-noise ratio of power sequences through energy threshold driven adaptive decomposition and provides high fidelity input for subsequent models;
2. Hierarchical feature abstraction: The four-layer pyramid structure of RBM (20-15-10-5 neurons) compresses the nonlinear dimension of the original data to 5-dimensional potential space, which retains the irradiance-power coupling relationship while suppressing the mode confusion caused by local shadows;
3. Rapid regression mechanism: The optimal number of nodes in the hidden layer of ELM (K=11) is determined by grid search, which reduces the analytical solution time of output weight to microsecond level under the premise of ensuring nonlinear mapping ability.

V. DISCUSSION AND CONCLUSION

The intermittency and volatility of distributed photovoltaic power generation pose challenges to grid stability. This study proposes a prediction method based on an improved Extreme Learning Machine, integrating Singular Spectrum Analysis (SSA) and Restricted Boltzmann Machine (RBM), to construct a "data denoising-feature learning-intelligent prediction" framework. SSA is used for multi-scale decomposition of power sequences to separate trends from noise, providing high-quality modeling data. A D-ELM hybrid architecture is constructed, where RBM extracts deep features unsupervised, and ELM combines dynamic weight adjustment with particle swarm optimization to enhance prediction accuracy. Experiments show that the normalized mean square error of the model for predictions at different time scales is 0.12kW and 0.08 kW, effectively handling nonlinear and non-stationary data. The study also develops a fully automated prediction system, integrating meteorological interfaces, real-time monitoring, and intelligent engines, supporting multi-site distributed deployment. Through MATLAB, adaptive learning and visual monitoring are achieved, enhancing the intelligent management level of photovoltaic power stations.

REFERENCES

[1]. Meng Panlei. Distributed Photovoltaic Power Dispatching System under the Internet of Things Architecture [J]. Electrical Engineering Technology, 2024, (22):54-56.
[2]. Lin Hongen, Zhao Erfeng, Liu Feng, et al. Prediction model for dam displacement combination considering subsequence characteristics [J]. Hydropower, 2025,51(03):113-118.
[3]. Hinton G E, Sejnowski T J. Learning and relearning in Boltzmann machines[M]//Parallel distributed processing: explorations in the microstructure of cognition. 1986: 282-317.
[4]. Liang Jiaye. Research on Security Defense Mechanisms for Intelligent Speech Systems Based on Singular Spectrum Analysis [D]. Disaster Prevention and Mitigation College, 2024.
[5]. Chen Danni. Continuous blood pressure estimation based on eigenplane singular spectrum analysis for noise reduction [D]. Guangdong University of Technology, 2023.
[6]. Zhang Zheng. Test and Analysis of Performance and Adaptability of New Open TBM [D]. Shijiazhuang Tiedao University, 2022.DOI:10.27334/d.cnki.gstdy.2022.000562.
[7]. Ding Runtian. Research on Parallel Operation of Photovoltaic Inverters Based on Virtual Synchronous Generator Control Method [D]. Xi'an University of Technology, 2023.
[8]. Lin Hongen, Zhao Erfeng, Liu Feng, et al. Prediction model of dam displacement combination considering subsequence characteristics [J]. Hydropower, 2025,51(03):113-118.
[9]. Jin Liyan, Chen Li, Fan Taiting, et al. Speech denoising algorithm based on singular spectrum analysis and Wiener filtering [J]. Computer Applications, 2015,35(08):2336-2340.
[10]. Liu Shaofeng, Zhao Shaoxiong, Zhou Wenshuo. Optimization and Performance Improvement of Intelligent O&M System Based on Machine Learning [J]. Digital Technology and Application, 2024,42(03):28-30.