Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Photovoltaic power forecasting using Bayesian optimized hybrid decomposition and a convolutional Transformer

Chenyu Qi
Shanghai Institute of Technology
Hua He
Shanghai University
Yuan Xue
xueyuan95@shu.edu.cn

Shanghai University

Article

Keywords: Photovoltaic power forecasting, Similar-day selection, Bayesian optimization, Variational mode decomposition, Multiscale permutation entropy, Convolutional Transformer

Posted Date: November 27th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11572606/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Photovoltaic power forecasting using Bayesian optimized hybrid decomposition and a convolutional Transformer

Chenyu Qi[1], Hua He[2], Yuan Xue[3,*]

[1] Faculty of Management, Shanghai Institute of Technology, Shanghai 201418, China

[2] School of Management, Shanghai University, Shanghai, China

[3] School of Intellectual Property, Shanghai University, Shanghai, China

*Correspondence: Yuan Xue (xueyuan95@shu.edu.cn).

Abstract

Accurate photovoltaic power forecasting supports the operation of electricity systems with variable renewable generation. We combine weighted similar-day selection, Bayesian optimized signal decomposition and a convolutional neural network–Transformer forecasting model. Pearson correlation analysis screens candidate features, and six complementary similarity measures identify historical days with comparable conditions. Bayesian optimization tunes variational mode decomposition, while multiscale permutation entropy identifies complex components for further processing by complete ensemble empirical mode decomposition with adaptive noise. The resulting components are used to train recurrent, attention-based and convolutional hybrid forecasting models. Evaluation uses the SOLETE photovoltaic dataset from Denmark, with 15-minute sampling and a target period of 24–30 June 2019. The convolutional Transformer achieves a mean absolute error of 0.4761, a root mean square error of 0.6582 and a coefficient of determination of 0.8888. Hybrid decomposition reduces its mean absolute error by 10.3% relative to variational mode decomposition alone and by 30.3% relative to adaptive-noise ensemble decomposition alone. Among the six forecasting models, the convolutional Transformer has the lowest root mean square error, whereas the convolutional gated recurrent unit has the lowest mean absolute error. For this dataset and target week, these results support further evaluation of local feature extraction, attention and selective signal decomposition.

Keywords: Photovoltaic power forecasting; Similar-day selection; Bayesian optimization; Variational mode decomposition; Multiscale permutation entropy; Convolutional Transformer

Introduction

As a crucial form of renewable energy, photovoltaic (PV) power generation has been widely applied in global energy production. In particular, in joint efforts to address global climate change and reduce greenhouse gas emissions, PV systems play an indispensable role. With the continuous innovation of PV technologies, the cost of PV power generation has decreased substantially, and various PV systems have penetrated almost all fields of energy production [1]–[4]. However, solar energy resources are highly susceptible to weather conditions, seasonal variations and diurnal cycles, resulting in prominent uncertainty in PV power output. This creates substantial challenges for the stable operation and dispatching of power systems. Against the background of high-penetration renewable energy integration, accurate and reliable PV power forecasting has therefore become an urgent and critical issue to address [5]–[7].

Numerous scholars have developed various models for PV power forecasting, which can be broadly classified into three categories. The first category refers to physical forecasting methods. Such approaches establish prediction models on the basis of detailed meteorological and environmental data. Although physical methods do not rely on massive amounts of historical power generation data, they exhibit poor anti-interference capabilities and fail to guarantee stable model accuracy [8]–[10]. The second category corresponds to classical statistical methods. These methods feature simple algorithm implementation and automated operational processes. Nevertheless, they require sufficient historical data for parameter calibration, along with inflexible threshold settings and poor adaptive adjustment capability [11], [12]. In contrast, the third category refers to machine learning methods, especially deep learning (DL) techniques. By simulating the structural characteristics of human neural networks, such methods can effectively mine the inherent nonlinear features hidden in PV power datasets. They overcome the inherent limitations of physical and traditional statistical approaches, thereby achieving a substantial improvement in forecasting accuracy [13]–[17]. Therefore, current research focuses predominantly on machine learning-based forecasting frameworks integrated with similar day selection, signal decomposition, and DL techniques. Such hybrid strategies have emerged as prominent research hotspots worldwide to further increase the accuracy of PV power forecasting.

In DL-based PV power forecasting, accurate and efficient selection of similar samples serves as a fundamental prerequisite. By constructing homogeneous daily datasets to narrow the range of candidate samples, the interference caused by dates with greatly different meteorological conditions from the target day can be effectively reduced, which provides a reliable basis for high-precision PV power prediction. For example, Zhan *et al*. proposed a similar day screening method that combines grey correlation analysis and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) decomposition and established a hybrid neural network model to realize multiscenario short-term PV power forecasting. Nevertheless, the low specificity of existing similar day selection methods limits the effective capture of the inherent long-term variability and physical monotonicity of PV systems, thereby resulting in degraded forecasting performance under volatile weather conditions [18]. Yang *et al*. developed a dual-dimensional screening method that combines the Euclidean distance and grey relational analysis for similar day selection. On this basis, an ensemble learning algorithm was adopted to weight and fuse multiple sets of prediction results to achieve short-term PV power forecasting. However, as global similarity measurement indicators, both ED and GRA fail to capture the local feature similarity of time series data effectively [19]. To address the challenges induced by temporal resolution variations in similar day analyses, Huang *et al*. proposed a method that integrates dynamic time warping (DTW) with a convolutional long short-term memory network to enhance forecasting performance. Nevertheless, excessive distortion generated by DTW may trigger unreasonable matching results, which further causes deviations between training data and prediction targets [20]. Xiang *et al*. proposed a Mahalanobis distance-based sample similarity measurement method

to quantify the covariance structural differences among multidimensional meteorological variables. Combined with partition entropy to optimize the clustering process, a cascaded data-driven forecasting framework was established for PV power output prediction. Nevertheless, when highly nonlinear relationships exist between meteorological factors and power output, the MD is insufficient to fully capture the similarity under such nonlinear structures, thereby resulting in inaccurate similar-day selection [21]. Dai *et al*. screened input variables highly correlated with PV power by utilizing the Spearman correlation coefficient and optimized the input structure through clustering to improve the forecasting performance from the perspectives of feature correlation and deep time series modeling. However, feature selection relying merely on the Spearman correlation coefficient cannot capture nonmonotonic strong correlations while easily neglecting multivariate interactions and lacking causal inference capability [22].

Most of the above studies accomplish similar-day screening on the basis of data correlation analysis. To comprehensively measure multi-scale feature similarity within time series, this study develops a comprehensive similarity evaluation method through the weighted combination of the Euclidean distance, grey relational degree, cosine similarity, DTW, Mahalanobis distance and Spearman correlation coefficient for similar day selection. This strategy enables the complex relationships among similar samples to be captured in a more comprehensive and flexible manner. In practical scenarios, meteorological and power load series generally exhibit non-stationary, multi-scale and strongly nonlinear characteristics, accompanied by severe random noise interference. Traditional linear analytical methods such as the Fourier transform assume signal stationarity, which limits their ability to extract inherent evolutionary rules of time series. Under such circumstances, modal decomposition techniques have become core tools for processing non-stationary signals. These methods adaptively decompose complex original signals into a set of intrinsic mode functions (IMFs) with clear physical interpretations, which effectively separate fluctuation components at different scales and provide a reliable foundation for subsequent feature similarity calculations on the basis of weighted distance integration. Among various decomposition algorithms, variational mode decomposition (VMD) has been widely applied in signal processing because of its superior anti-mode aliasing performance and solid mathematical theoretical basis [23]–[26]. Nevertheless, the decomposition performance of VMD is highly dependent on the selection of critical hyperparameters, including the penalty factor α and mode number K. Traditional empirical parameter configuration strategies fail to adapt to dynamic signal variations and may lead to over-decomposition or under-decomposition, which directly deteriorates the accuracy of subsequent analysis. In recent years, intelligent optimization algorithms have been introduced to address this limitation. For example, Zhao *et al*. optimized VMD parameters via the whale optimization algorithm (WOA) to obtain decomposed components containing optimal feature information of PV power time series. The optimized subsequences were further imported into the forecasting model to increase the accuracy of short-term PV power prediction. However, the WOA still has inherent drawbacks, such as local optimum trapping, limited convergence speed and insufficient parameter adaptability [27]. Cui et al. optimized VMD parameters via the artificial hummingbird algorithm (AHA) and established an AHA-VMD model. Combined with multi-scale permutation entropy (MPE) for sequence reconstruction, the effectiveness of the proposed method was validated via measured data from multiple PV power stations, which verified the superior prediction accuracy and stability. Nevertheless, the convergence speed of the AHA and the quality of its optimal solutions are susceptible to its inherent hyperparameters and the generation strategy of the initial population [28]. Additionally, Zhang *et al*. adopted the grey wolf optimization (GWO) algorithm to automatically tune the critical parameters of VMD. The original wind power data were decomposed into a series of IMFs, and the fuzzy entropy of each IMF was subsequently calculated. All subsequences were further reconstructed into trend components and stochastic components, which were fed into the forecasting model for training and prediction. This approach effectively improved the prediction accuracy of the wind power series.

Nonetheless, the optimization performance and computational efficiency of GWO are highly restricted by its inherent hyperparameters, such as population size and maximum iteration number [29].

Although intelligent optimization algorithms can mitigate the parameter sensitivity of VMD to a certain extent, the parameter optimization process still suffers from high computational cost and local optimum trapping. These inherent shortcomings make fundamentally eliminating the adverse effects of inappropriate parameter selection on decomposition quality difficult. To address this issue, several scholars have shifted their focus to adaptive decomposition methods that require no prespecified parameters. Among these approaches, CEEMDAN has emerged as a vital alternative and complementary technique to VMD owing to its outstanding adaptive decomposition capability [30]–[32]. The core advantage of CEEMDAN lies in its ability to adaptively decompose complex original signals into a series of IMFs with distinct feature scales and negligible noise interference, which effectively avoids decomposition distortion caused by improper parameter settings in VMD. For example, Li *et al*. combined CEEMDAN with a long short-term memory (LSTM) network. The original time series was decomposed into subsequences with different frequency characteristics through CEEMDAN, and individual prediction was conducted for each component, thereby prominently improving the accuracy of ultrashort-term power load forecasting. Nevertheless, excessive IMFs generated by CEEMDAN greatly increase the computational burden and model complexity, restricting its practical engineering application [33]. Liang et al. adopted CEEMDAN to forecast each reconstructed subsequence and obtain the final prediction result through sequence superposition, which effectively enhanced the forecasting accuracy of the PV output. However, the independent modeling of all decomposed subsequences based solely on CEEMDAN inevitably leads to a time-consuming forecasting process and low computational efficiency [34].

Overall, single-signal decomposition methods still exhibit inherent bottlenecks when coping with actual highly complex signals with mixed multi-feature components. Although VMD can effectively separate high-frequency components and suppress noise owing to its non-recursive decomposition mode, its high sensitivity to boundary conditions tends to cause decomposition distortion when processing complex modes under strong noise interference. Moreover, CEEMDAN refines local features through a noise-assisted mechanism; however, it cannot eliminate endpoint effects and mode aliasing thoroughly. Accordingly, a clear trade-off exists between robustness and completeness for single decomposition paradigms, which demonstrates the necessity of establishing a hybrid and complementary decomposition strategy. In summary, this paper proposes a systematic hybrid decomposition and forecasting framework on the basis of previous studies. First, a comprehensive similarity indicator integrating six similarity metrics is adopted to screen similar days, with the aim of reducing mismatch between historical samples and target conditions. Second, Bayesian optimization is introduced to intelligently tune the key hyperparameters of VMD, improve the rationality of the parameter configuration and the stability of the decomposition results, and effectively reduce decomposition errors caused by setting empirical parameters. The MPE is subsequently utilized to quantitatively evaluate the complexity of each modal component decomposed via Bayesian-optimized VMD. According to the complexity characteristics of different components, the corresponding subsequences are dynamically assigned to CEEMDAN for secondary refined decomposition, forming a hierarchical processing mechanism of "coarse decomposition to fine decomposition". This design combines the advantages of VMD in high-frequency component separation and noise suppression, as well as the strengths of CEEMDAN in adaptively capturing local features and alleviating mode aliasing, with the aim of improving feature extraction and modal separation for the original PV power series. Finally, individual forecasting models are constructed for each decomposed subsequence, and the prediction outputs of all the submodels are integrated to increase the overall accuracy and reliability of PV power forecasting. As a core link for performance improvement, intelligent forecasting algorithms can effectively optimize the precision of prediction results [35]–[38]. This study compares multiple neural network models within the proposed forecasting

framework. On this basis, a hybrid decomposition and forecasting system driven by comprehensive similarity is constructed, whose performance is evaluated on the selected dataset and target period.

Methods

Framework

The proposed framework combines feature identification, similar-day selection, hybrid modal decomposition and forecasting model comparison for non-stationary photovoltaic power series. The preprocessing procedures draw on previous photovoltaic forecasting studies [39], [40]. The workflow is illustrated in Fig. 1, and its individual steps are described below.

Feature factor screening

To examine the influencing factors of PV output, this paper first establishes a physics-informed data array by integrating and extending the King PV model and the Rincon–Pombo thermodynamic model [41]. To eliminate dimensional differences and quantify feature correlations, Z score standardization and Pearson correlation analysis are subsequently performed on the meteorological factors, physical parameters, and PV power data. Redundant features with low measured correlations with PV output are ultimately removed to screen for key influencing factors.

Similar Day Selection

To improve the quality of reference data for the forecast day, this paper calculates six metrics between the time series of historical days and the target day, namely, the Euclidean distance, grey relational degree, cosine similarity, DTW, Mahalanobis distance and Spearman correlation coefficient. After normalization, weighted fusion of these indicators is conducted to establish a comprehensive similarity evaluation method. On this basis, historical days with highly consistent meteorological and environmental conditions with the target day are accurately selected, providing highly representative time series samples for subsequent model training.

Hybrid Decomposition Strategy

To address the limitation that single modal decomposition methods cannot balance decomposition efficiency and complex signal processing capability, this study adopts a hybrid decomposition strategy. First, Bayesian optimization is employed to intelligently optimize the key parameters of VMD, including the penalty factor and modal number, to decompose the original series into multiple subsequences. Subsequently, MPE is introduced to quantify signal complexity, and CEEMDAN is utilized for secondary refined decomposition of high-complexity modes. Finally, frequency similarity grouping is realized on the basis of entropy evaluation and covariance structure analysis, and refined components are reconstructed to generate new subsequences. The proposed strategy can effectively suppress mode aliasing, reduce the nonlinearity and non-stationarity of time series, and provide easily modeled temporal components for subsequent prediction.

Prediction Model Construction

To model local features and longer temporal dependencies within time series data, three basic neural network models, namely, LSTM, a Gated Recurrent Unit (GRU) and a Transformer, are established in this paper. Moreover, a Convolutional Neural Network (CNN) is introduced to capture local features, thereby constructing three combined neural network frameworks: CNN-LSTM, CNN-GRU, and CNN-Transformer. The CNN can efficiently mine the local fluctuation characteristics of meteorological data, whereas the LSTM, GRU and Transformer methods have advantages in time-dependent modeling, computational efficiency and global correlation capture of long sequences, respectively. The integration of these networks enables the collaborative modeling of spatial and temporal features. Afterwards, the

subsequences obtained via hybrid decomposition are transformed into feature matrices and fed into the above models for training and forecasting.

Model Performance Evaluation

To comprehensively compare the forecasting performance of different models, multiple evaluation metrics are adopted in this study, including the mean absolute error (MAE), root mean square error (RMSE), relative squared error (RSE), and coefficient of determination (R^2). On this basis, a comprehensive performance assessment is conducted for single neural networks (LSTM, GRU, Transformer) and CNN-based hybrid neural networks under three decomposition schemes, namely, CEEMDAN, VMD, and the proposed hybrid decomposition method. The comparisons identify the best-performing configurations among the tested models for the selected dataset and target period.

Data preprocessing

Z score normalization

Features have different units and numerical scales, which can bias model training toward variables with larger magnitudes. Z score normalization centres each feature using its sample mean and scales it using its sample standard deviation. The standardized values have zero mean and unit standard deviation; this transformation does not, by itself, imply a normal distribution. The normalization is defined as follows:

$$z=\frac{x-\mu}{\sigma} \tag{1}$$

where x represents the original data, μ denotes the sample mean, σ is the sample standard deviation, and z refers to the normalized data. Z score normalization can effectively eliminate the dimensional differences among various variables and provide a consistent numerical scale for model inputs. Moreover, on the premise of retaining the distribution characteristics of the original data, this method enhances the comparability between different features, which is applicable to the subsequent similarity measurement and machine learning model training.

Pearson correlation coefficient

To measure the degree of linear correlation between different variables quantitatively, the PCC is introduced in this paper as a similarity measurement indicator [42]. The PCC can reflect the strength and direction of the linear correlation between two variables, and its value ranges from -1 to 1. The corresponding calculation formula is presented as follows:

$$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{2}$$

where x_i and y_i represent the values of two variables at the i-th sample point, $\bar{x}$ and $\bar{y}$ are the corresponding variable means, and n denotes the total number of samples. When r=1, the two variables are completely positively correlated; when r=-1, they are completely negatively correlated; and when r=0, there is no linear correlation between them. By calculating the PCCs between the PV output and each candidate feature factor (including irradiance, ambient temperature, wind speed, humidity, PV module temperature, and cell temperature), features with strong linear associations on PV power generation can be screened out, while factors with low correlations can be eliminated. This procedure provides a reduced feature set for the subsequent forecasting models.

Comprehensive similarity weighting

Six similarity indicators are combined to evaluate historical days relative to a target day. Each indicator is normalized so that larger values represent greater similarity, and the normalized scores are combined

by weighted summation. The analytic hierarchy process (AHP) provides a structured framework for expressing pairwise preferences and checking their consistency [43]–[45].

Hierarchical structure

The core of the AHP lies in decomposing complex multi-criteria decision-making problems into a hierarchical structure and realizing the organic combination of subjective judgment and objective calculation through hierarchical quantification. In accordance with the research requirements of time series comprehensive similarity evaluation, a three-level hierarchical framework consisting of a target layer, criterion layer and indicator layer is constructed in this paper to systematically integrate multidimensional similarity indicators and complete weight assignment.

Target Layer (G)

The target layer serves as the highest level of the hierarchical structure, which defines the core objective of the evaluation, namely, the optimal evaluation of comprehensive time series similarity. Centering on the essential demand of time series similarity measurement, this objective integrates multidimensional evaluation indicators to establish a comprehensive assessment system that takes geometric morphology, statistical correlation and directional characteristics into account. This approach provides a scientific quantitative basis for time series matching, prediction accuracy verification, anomaly detection and other related research.

Criterion Layer

The criterion layer decomposes the overall objective from multiple dimensions. On the basis of the essential characteristics of time series similarity, comprehensive similarity is divided into three independent and logically complementary primary criteria to overcome the limitations of single-index measurement:

C_1: Geometric distance similarity. It characterizes the absolute deviation of time series in the numerical space and reflects the overall numerical fitting accuracy, serving as the fundamental dimension for similarity measurement.

C_2: Statistical correlation similarity. It quantifies the correlation degree of time series in terms of statistical rules and describes the consistency of variation trends, which is applicable to nonlinear and non-normal time series data.

C_3: Morphological direction similarity. It represents the matching degree of a time series in various directions and reflects the consistency of fluctuation patterns. This criterion weakens the influence of absolute numerical errors and highlights trend characteristics.

Indicator Layer (I)

The indicator layer consists of concrete measurable metrics corresponding to the criterion layer. In accordance with the three primary criteria, six classic and highly adaptive time series similarity indicators are selected as secondary evaluation indices. The corresponding relationship and connotation of each indicator are as follows:

I_1: Euclidean distance. The geometric distance similarity criterion is among the most widely adopted numerical distance metrics. By calculating the square root of the sum of the squared differences at the corresponding points between two sequences, the absolute numerical deviation of the time series can be quantified. The calculation formula is expressed as follows (where n denotes the sample size):

$$d_{ED}(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{3}$$

where $X=(x_1,x_2,\ldots,x_n)$ and $Y=(y_1,y_2,\ldots,y_n)$ denote two time series of equal length. As a negative indicator, a smaller value indicates greater similarity between sequences.

I_2: Grey Relational Analysis. Owing to the statistical correlation similarity criterion, this indicator is derived from grey system theory. It evaluates the correlation degree of sequences by quantifying the geometric similarity of sequence curves, with no strict restrictions on sample size or data distribution. Its formula is as follows:

$$r(X,Y)=\frac{1}{n}\sum_{i=1}^{n}\xi_i \tag{4}$$

where ξ_i represents the correlation coefficient of the i-th sampling point. As a positive indicator, a larger value corresponds to stronger sequence similarity.

I_3: Cosine similarity. The morphological direction similarity criterion measures the directional consistency of sequence variations by calculating the cosine value of the spatial angle between two vectors. This metric weakens the interference of absolute numerical values and highlights trend characteristics. The formula is as follows:

$$\cos\theta=\frac{X\cdot Y}{\|X\|\|Y\|}=\frac{\sum_{i=1}^{n}x_i y_i}{\sqrt{\sum_{i=1}^{n}x_i^2}\sqrt{\sum_{i=1}^{n}y_i^2}} \tag{5}$$

It is a positive indicator, and a value closer to 1 signifies greater similarity in sequence variation trends.

I_4: DTW. Under the geometric distance similarity criterion, nonlinear time alignment is realized to address the time axis offset and inconsistent sequence length. It is capable of effectively measuring the morphological similarity of unaligned time series. The expression is:

$$D_i^{(DTW)}=\min\sum_{(p,q)\in W}(x_p-x_{i,q})^2 \tag{6}$$

where W represents the optimal matching path. As a negative indicator, a lower DTW value means a higher degree of sequence similarity.

I_5: Mahalanobis distance (MD). The covariance structure of the data is considered in the geometric distance similarity criterion, which eliminates the influence of variable correlation and dimensional differences. It is highly applicable to multidimensional time series with correlated features. The formula is as follows:

$$d_{MD}(X,Y)=\sqrt{(X-Y)^T\Sigma^{-1}(X-Y)} \tag{7}$$

where Σ denotes the covariance matrix of the time series. It is a negative indicator, and a smaller value indicates better similarity.

I_6: Spearman's rank correlation coefficient. Owing to the statistical correlation similarity criterion, it is a nonparametric rank correlation metric. It quantifies the consistency of sequence variation trends on the basis of rank correlation, which is insensitive to data distribution and outliers. The corresponding formula is as follows:

$$\rho_s=1-\frac{6\sum_{i=1}^{n}d_i^2}{n(n^2-1)} \tag{8}$$

where d_i is the rank difference of the corresponding elements. As a positive indicator, a value closer to 1 reflects greater consistency in sequence variation trends.

Pairwise comparison matrix

Adopting the 1–9 scale method to conduct pairwise comparisons among indicators (see Supplementary Table S1 for the values and corresponding definitions of the 1–9 scale), the judgment matrix is further constructed:

$$A=(a_{ij})_{n\times n} \tag{9}$$

It satisfies the following conditions:

$$a_{ij}=\begin{cases} 1 & i=j \\ \frac{1}{a_{ji}} & i\neq j \end{cases} \tag{10}$$

where a_{ij} represents the importance of indicator i relative to indicator j and where n denotes the number of indicator dimensions. The corresponding matrix form is expressed as:

$$A=\begin{bmatrix} 1 & a_{12} & \cdots & a_{1j} \\ \frac{1}{a_{12}} & 1 & \cdots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{a_{1j}} & \frac{1}{a_{2j}} & \cdots & 1 \end{bmatrix} \tag{11}$$

Weight vector calculation

By solving the maximum eigenvalue of the judgment matrix and its corresponding eigenvector, the weight coefficients can be obtained:

$$Aw=\lambda_{max}w \tag{12}$$

and the eigenvector is further processed via normalization:

$$w_i=\frac{v_i}{\sum_{i=1}^{n} v_i},\ i=1,2,\ldots,n \tag{13}$$

where λ_{max} denotes the maximum eigenvalue of the judgment matrix and w_i denotes the weight of the i-th indicator and where v_i represents the component of the eigenvector.

Consistency test

Because the pairwise comparison matrix reflects subjective judgements, a consistency test is used to assess the internal coherence of those judgements.

Consistency Index (CI)

$$CI=\frac{\lambda_{max}-n}{n-1} \tag{14}$$

Consistency ratio (CR)

$$CR=\frac{CI}{RI} \tag{15}$$

where RI refers to the random consistency index, which is determined by the order of the judgment matrix.

Judgment criterion

If $CR<0.1$, the judgment matrix possesses acceptable consistency; otherwise, it is necessary to readjust the matrix A.

Weight vector

The final weight vector is obtained as follows:

$$W=(w_1,w_2,w_3,w_4,...,w_n) \tag{16}$$

and it satisfies:

$$\sum_{i=1}^{n} w_i = 1 \tag{17}$$

Application of similarity weights

The obtained weights are introduced into the comprehensive similarity model:

$$Sim_i = \sum_{j=1}^{n} w_j \cdot S_i^{(j)} \tag{18}$$

where $S_i^{(j)}$ is the normalized result of the *j*-th indicator.

Normalization of similarity indicators

In the multi-index similarity evaluation system, different similarity measurement indicators differ greatly in physical meaning, dimension scale and numerical range. For distance-based indicators (e.g., Euclidean distance, Mahalanobis distance, and DTW distance), a smaller value indicates greater similarity between sequences. In contrast, for similarity-based indicators (e.g., the Spearman rank correlation coefficient, cosine similarity, grey correlation degree), a larger value corresponds to stronger sequence similarity. To eliminate the interference of dimensional differences on comprehensive evaluation results and ensure the comparability of various indicators, separate normalization transformations are performed for the two types of indicators. All the indicators are uniformly mapped to the interval [0, 1], and a larger normalized value represents greater similarity.

Distance indicators

For distance-based indicators (Euclidean distance, Mahalanobis distance, DTW), the original values are negatively correlated with similarity. Therefore, the minimized range normalization method is adopted to inversely map distance values to a unified interval through linear transformation. The formula is expressed as follows:

$$S_i^{(j)} = \frac{D_{max}^{(j)} - D_i^{(j)}}{D_{max}^{(j)} - D_{min}^{(j)}} \tag{19}$$

where *j* denotes the index dimension, representing the *j*-th distance-based similarity indicator; *i* is the sample index, corresponding to the *i*-th sample to be evaluated; $D_i^{(j)}$ refers to the original calculated value of the *i*-th sample under the *j*-th distance indicator; $D_{max}^{(j)}$ and $D_{min}^{(j)}$ are the maximum and minimum values of all samples for the *j*-th distance indicator, respectively; and $S_i^{(j)}$ represents the normalized result of the *i*-th sample on the *j*-th distance indicator, which ranges within [0,1], where a larger value indicates a greater similarity between samples.

Similarity indicators

For similarity-based indicators, their original values are positively correlated with similarity. Accordingly, the maximization range normalization method is applied to positively map the similarity values to the interval [0,1] via linear transformation, and the formula is as follows:

$$S_i^{(j)} = \frac{S_i^{(j)} - S_{min}^{(j)}}{S_{max}^{(j)} - S_{min}^{(j)}} \tag{20}$$

where j is the index dimension denoting the j-th similarity-type indicator and where i refers to the sample index representing the i-th sample to be evaluated; $S_i^{(j)}$ is the original calculated value of the i-th sample under the j-th similarity indicator; $S_{\max}^{(j)}$ and $S_{\min}^{(j)}$ are the maximum and minimum values of all samples corresponding to the j-th similarity indicator, respectively; and the normalized result of the i-th sample for the j-th similarity indicator is constrained within the range of [0,1], and a larger value implies greater similarity between samples.

Through the above two normalization transformations, the dimensionality reduction and homogenization of distance-based and similarity-based indicators can be realized. All indicators are uniformly mapped to the interval [0,1], and their values are positively correlated with the degree of similarity, which lays a foundation for the subsequent multi-index comprehensive similarity evaluation.

Bayesian optimization of variational mode decomposition

Variational mode decomposition

VMD is an adaptive non-recursive signal decomposition method [46]. It can realize adaptive signal decomposition and modal component extraction in the frequency domain and effectively suppress mode aliasing and endpoint effects. Its decomposition performance is highly dependent on the reasonable configuration of two core parameters, namely, the mode number K and the penalty factor α. Traditional manual parameter tuning suffers from strong subjectivity, low efficiency and insufficient parameter optimization accuracy, which makes it difficult to meet the decomposition requirements of complex signals. The core principle of VMD is to construct a constrained variational problem and adaptively decompose the original signal into K band-limited modal components. The constrained variational problem of VMD is first established to minimize the sum of the squared bandwidths of each modal component, with the constraint that the sum of all modal components is equal to the original signal. The corresponding formula is expressed as follows:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\} \quad \text{s.t.} \quad \sum_{k=1}^{K} u_k(t) = f(t) \tag{21}$$

where u_k represents the set of K decomposed modal components, ω_k denotes the central frequency set of each modal component, $\delta(t)$ is the Dirac function, * refers to the convolution operation, $f(t)$ is the original signal, and ∂_t stands for the partial derivative with respect to time t.

To convert the constrained variational problem into an unconstrained problem, a quadratic penalty term and Lagrange multiplier λ are introduced to construct the augmented Lagrangian function.

$$L(\{u_k\},\{\omega_k\},\lambda) = \alpha \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta + \frac{j}{\pi t} \right) * u_k \right] e^{-j\omega_k t} \right\|_2^2 + \left\| f - \sum_{k=1}^{K} u_k \right\|_2^2 + \left\langle \lambda, f - \sum_{k=1}^{K} u_k \right\rangle \tag{22}$$

where α denotes the penalty factor, which controls the bandwidth constraint strength of each mode, and $\langle \cdot \rangle$ represents the inner product operation. The alternating direction method of multipliers is adopted to iteratively update the modal components, central frequencies and Lagrange multipliers to solve the variational problem. The core update formulas for the modal components and central frequencies are as follows:

Modal component updating (frequency-domain form, where $\hat{\cdot}$ denotes the Fourier transform):

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i \neq k} \hat{u}_i^n(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k^n)^2} \tag{23}$$

Central frequency updating:

$$\omega_k^{n+1}=\frac{\int_0^\infty \omega\left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega} \tag{24}$$

Adaptive parameter optimization

To address the deficiencies of traditional manual parameter tuning, this study proposes an adaptive parameter optimization method for VMD based on Bayesian optimization. By constructing an intelligent parameter search space that fits signal characteristics and designing a dual-objective optimization function considering both reconstruction accuracy and mode separation degree, the proposed method searches for suitable values of the mode number K and penalty factor α. It provides an adaptive parameter configuration scheme for VMD decomposition (the optimization process is illustrated in Supplementary Fig. S1). Traditional VMD parameter optimization mostly adopts a fixed search interval strategy with obvious limitations. An excessively wide search interval increases the number of invalid search iterations and reduces the optimization efficiency, whereas an overly narrow interval tends to omit optimal parameters and leads to optimization failure. The convergence efficiency of Bayesian optimization is closely related to the rationality of the search space. Existing studies have demonstrated that Bayesian optimization presents higher sample efficiency and convergence stability under low-dimensional and moderately constrained search intervals [47], [48]. When the search space is too wide or the dimension is excessively high, the fitting complexity of the surrogate model increases substantially, thereby degrading the optimization efficiency [49]. Therefore, establishing an adaptive parameter search space on the basis of signal characteristics is the key to balancing the global search capability and optimization efficiency. In this study, a customized intelligent parameter search space is constructed with inherent signal features. The search ranges of the mode number K and penalty factor α are dynamically adjusted according to key characteristics such as the signal length and fluctuation complexity [50], [51]. The specific implementation is described as follows:

Mode number constraint on the basis of the signal length

The mode number K directly affects the frequency decomposition granularity of VMD. An excessively small K will cause frequency aliasing, whereas an overly large K leads to over decomposition and spurious modes. According to the principle of signal frequency resolution, the sample length directly affects the frequency discrimination capability of a signal. For short samples with limited frequency resolution, an excessive number of modes should be avoided. As the sample length increases, more frequency components can be distinguished, and the upper limit of the mode number can be appropriately increased. Moreover, statistical stability constraints are introduced to reduce the risk of overfitting. On this basis, a hierarchical search interval for the mode number is established, which meets the requirement of frequency discrimination and realizes a controllable scale of the search space.

Dynamic Upper Limit Adjustment Based on Fluctuation Complexity

The signal standard deviation serves as a core indicator reflecting signal fluctuation complexity; a more drastically fluctuating signal corresponds to higher complexity. The penalty factor α directly governs the bandwidth constraint strength of each VMD modal component and acts as a critical parameter for suppressing mode aliasing. For signals with high fluctuation and complexity, an insufficient penalty factor easily results in band overlap and mode aliasing. Accordingly, the search upper bound of the penalty factor is dynamically adjusted according to the signal fluctuation complexity. This measure strengthens bandwidth constraints to avoid modal cross-coupling and simultaneously prevents signal distortion caused by excessive parameter shrinkage. In addition, global parameter constraints are formulated to prevent VMD failure induced by out-of-bound parameters.

Design of the Optimization Objective Function

To evaluate VMD configurations, the optimization objective combines reconstruction error and inter-mode correlation. The objective penalizes inaccurate reconstruction and correlated components; a smaller value is preferred. It is defined as follows:

$$J(K,\alpha)=\lambda_1 MSE(x,\hat{x})+\lambda_2 \sum_{1\le i<j\le K}\left|\rho_{ij}\right| \tag{25}$$

The reconstruction error term $MSE(x,\hat{x})$ is adopted to measure the reconstruction accuracy after decomposition and characterize the degree of fit between the reconstructed signal and the original signal. Here, $\hat{x}(t)=\sum_{k=1}^{K} u_k(t)$ denotes the reconstructed signal obtained via VMD. The mode separation penalty term $\sum_{1\le i<j\le K}|\rho_{ij}|$ quantifies the degree of overlap among IMF components, where ρ_{ij} represents the PCC between the i-th and j-th IMF components. A smaller penalty term indicates a higher degree of mode separation. The weight coefficients λ_1 and λ_2 are utilized to balance the two optimization objectives of reconstruction accuracy and mode separation. Through reasonable weight allocation, the coordinated optimization of decomposition results in terms of reconstruction accuracy and mode orthogonality, which adapts well to the signal decomposition requirements of this study.

Multiscale permutation entropy

The MPE is an effective tool for characterizing the nonlinear complex dynamic characteristics of signals. Its core principle is to perform multi-scale coarse-grained transformation on the original signal and calculate the permutation entropy (PE) of the sequence at different time scales to realize multidimensional quantitative analysis of signal complexity [52]–[54]. The quantitative results of MPE can directly reflect the complexity of a signal at corresponding scales. A larger MPE value indicates greater complexity and stronger randomness of the sequence at that scale; conversely, a smaller MPE value signifies more prominent regularity and certainty of the sequence.

The calculation process of MPE mainly includes three steps: coarse-grained processing, PE calculation and multi-scale fusion. The core formulas are presented as follows:

Coarse-grained Processing of Time Series

For the original time series $x=\{x_1,x_2,\ldots,x_N\}$, coarse-grained processing is carried out under the scale factor τ to obtain the multi-scale coarse-grained sequence $y^{(\tau)}$, and the corresponding formula is expressed as:

$$y_j^{(\tau)}=\frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i,\ j=1,2,\ldots,\left[\frac{N}{\tau}\right] \tag{26}$$

Phase space reconstruction of a coarse-grained sequence

Phase space reconstruction is performed on the coarse-grained sequence of each scale to acquire the phase point:

$$Y_i=\{y_i^{(\tau)},y_{i+\delta}^{(\tau)},\ldots,y_{i+(m-1)\delta}^{(\tau)}\} \tag{27}$$

where m is the embedding dimension and where δ represents the delay time. The probability distribution $P=\{p_1,p_2,\ldots,p_K\}$ is statistically calculated according to the permutation patterns of all phase points, and the single-scale PE is further defined as:

$$H_{pe}(m)=-\sum_{i=1}^{K} p_i \ln p_i \tag{28}$$

The PE values at all scales are averaged to obtain the MPE. Its normalized form constrains the entropy value within the range of $0\le H_{npe}\le 1$, which facilitates the quantitative discrimination of signal complexity. The corresponding normalization formula is as follows:

$$H_{npe}=\frac{H_{pe}(m)}{\ln(m!)} \tag{29}$$

In the signal preprocessing flow of this study, MPE is adopted to realize the adaptive optimization of the VMD results. Specifically, the MPE of each modal component obtained via VMD decomposition is calculated one by one. A discrimination criterion is established on the basis of the normalized characteristics of PE. After normalization, the value of PE ranges from 0 to 1. A PE value close to 0 indicates a highly regular sequence, whereas a value close to 1 represents an approximately random sequence. The intermediate PE value corresponds to the critical transition state from regularity to randomness.

The decomposition decision is based on component complexity. VMD components whose average MPE exceeds the threshold are selected for further CEEMDAN decomposition and grouping. Components below the threshold are retained without secondary decomposition. This rule is intended to concentrate further processing on complex components while limiting unnecessary decomposition of more regular signals. Its contribution to forecast performance requires evaluation through comparisons of the threshold and grouping choices.

Complete ensemble empirical mode decomposition with adaptive noise

Empirical mode decomposition (EMD) is a classic adaptive signal decomposition method. However, it suffers from inherent defects such as mode aliasing and end effects when processing non-stationary and nonlinear signals. Ensemble empirical mode decomposition (EEMD) suppresses mode aliasing by adding Gaussian white noise to the original signal with a uniform frequency distribution; however, it still suffers from residual noise and low computational efficiency.

CEEMDAN extends EEMD by introducing adaptively scaled noise at successive decomposition stages and averaging the resulting components [55]. The procedure is designed to improve decomposition completeness and reduce residual noise. These properties alone do not establish the computational cost of the complete forecasting pipeline, which requires separate benchmarking.

Decomposition procedure

Let the original signal be $x(t)$; then, the complete decomposition process of CEEMDAN can be fully described by the following mathematical formulas:

Initialization and noise addition

Let $E_k(\cdot)$ denote the k-th IMF obtained via EMD decomposition, $\omega^i(t)$ represent the i-th group of Gaussian white noise with zero mean and unit variance, and ε_0 be the initial noise amplitude coefficient. A total of I sets of Gaussian white noise are superimposed on the original signal $x(t)$ to construct I groups of noisy signals:

$$x^i(t)=x(t)+\varepsilon_0\omega^i(t), \quad i=1,2,\ldots,I \tag{30}$$

Extraction of the First-Order IMF Component

EMD decomposition is performed on each noisy signal $x^i(t)$ to extract the first-order IMF component $IMF_1^i(t)$. The average value of all the components is calculated to obtain the final first-order IMF component of CEEMDAN:

$$\overline{IMF_1}(t)=\frac{1}{I}\sum_{i=1}^{I} I\,MF_1^i(t)=\frac{1}{I}\sum_{i=1}^{I} E_1\left[x^i(t)\right] \tag{31}$$

Calculation of the first-order residual component

The first-order residual is acquired by subtracting the first-order IMF component from the original signal:

$$r_1(t)=x(t)-\overline{IMF_1}(t) \tag{32}$$

Iterative Extraction of Subsequent IMF Components

Gaussian white noise with an adaptively adjusted amplitude is added to the residual $r_1(t)$ to generate a new noisy residual sequence:

$$r_1^i(t)=r_1(t)+\varepsilon_1 E_1\left[\omega^i(t)\right], \quad i=1,2,\ldots,I \tag{33}$$

EMD decomposition is implemented on $r_1^i(t)$ to extract the first-order IMF component, and the average value is then calculated to obtain the final second-order IMF component:

$$\overline{IMF_2}(t)=\frac{1}{I}\sum_{i=1}^{I} E_1\left[r_1^i(t)\right] \tag{34}$$

Update the residual component:

$$r_2(t)=r_1(t)-\overline{IMF_2}(t) \tag{35}$$

The above iterative process is repeated until the residual sequence $r_k(t)$ meets the termination condition, i.e., the residual presents a monotonic function or the number of extreme points is less than 2. Ultimately, all the IMF components of each order and the final residual $r_K(t)$ are obtained.

Signal reconstruction formula

The original signal $x(t)$ can be completely reconstructed as the sum of all order IMF components and the final residual:

$$x(t)=\sum_{k=1}^{K}\overline{IMF_k}(t)+r_K(t) \tag{36}$$

where K denotes the total number of decomposed IMF components and where $r_K(t)$ is the final residual component that represents the trend term of the signal.

In this study, high-complexity modal components decomposed via VMD are further processed via CEEMDAN and reconstructed into more stationary subsequences via overlap processing. This approach effectively mitigates the interference of signal non-stationarity on subsequent feature extraction and provides fine-grained feature support for accurately selected similar days.

Evaluation metrics

Common model error evaluation indicators include the MAE, RMSE, RSE, and R^2. The corresponding calculation formulas and physical interpretations are presented below.

The MAE reflects the average level of absolute error between the predicted and actual values and exhibits favorable robustness to outliers. Its calculation formula is defined as:

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-\hat{y}_i| \tag{37}$$

where y_i denotes the actual value, $\hat{y}_i$ represents the predicted value, and n is the total number of samples. This indicator intuitively characterizes the average magnitude of prediction deviation with the same unit as the original data, facilitating practical engineering interpretation and quantitative analysis.

The RMSE restores the dimension to that of the original data. It possesses both error amplification characteristics and excellent interpretability:

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2} \tag{38}$$

RSE compares the model error with that of the benchmark model, which generally adopts the mean of actual values as a constant prediction. It is utilized to quantify the improvement degree of the proposed model relative to simple baseline methods:

$$\mathrm{RSE}=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{39}$$

where $\bar{y}$ refers to the sample mean of the true values. An RSE value less than 1 indicates that the established model outperforms the mean-value prediction, and a smaller RSE corresponds to a stronger explanatory capability of the model.

R^2 quantifies the proportion of data variation explained by the model, with a value range of (-∞,1]. A value closer to 1 indicates a higher goodness of fit.

$$\mathrm{R}^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{40}$$

In the model evaluation of this paper, the above four indicators are comprehensively adopted to construct a multidimensional evaluation system. Specifically, the MAE is employed to measure the absolute level of prediction errors; the RMSE serves as a core indicator integrating error amplification characteristics and dimensional consistency; and RSE and R^2 are utilized to assess the relative improvement of the proposed model over baseline schemes. For comparative experiments involving different decomposition strategies and forecasting models, quantitative comparisons based on these indicators are conducted to objectively evaluate the relative performance of each method in the PV power forecasting task.

Dataset and feature preparation

The study uses the SOLETE dataset, which includes measured meteorological variables and co-located photovoltaic power output in Denmark [56]. The selected records span 1 June 2018 to 30 June 2019, with daily observations from 05:00 to 20:00. Records at a 5-minute interval are resampled to a 15-minute resolution. The King photovoltaic performance model and the Rincón–Pombo thermodynamic model are used for feature expansion [41], followed by standardization and correlation-based screening. Feature definitions are provided in Table 1; the correlation plots are shown in Supplementary Figs. S2 and S3. The target period is the final week, 24–30 June 2019, for which historical similar days are selected for decomposition and forecasting.

Similarity weights and historical day selection

Trend matching capability indicators, including the grey correlation degree and DTW, are assigned relatively high importance in the pairwise comparison. The Mahalanobis distance accounts for multi-feature correlation, with an importance level close to that of DTW. The Euclidean distance and Spearman correlation coefficient serve as basic and supplementary indicators, respectively. Cosine similarity mainly characterizes the directional consistency of sequences and has relatively low importance. On this basis, the judgment matrix is established following the 1–9 scale method ($A=(a_{ij})$):

$$A=\begin{bmatrix} 1 & \frac{1}{2} & 2 & \frac{1}{3} & \frac{1}{3} & 1 \\ 2 & 1 & 3 & 1 & 1 & 2 \\ \frac{1}{2} & \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{2} \\ 3 & 1 & 3 & 1 & 1 & 2 \\ 3 & 1 & 3 & 1 & 1 & 2 \\ 1 & \frac{1}{2} & 2 & \frac{1}{2} & \frac{1}{2} & 1 \end{bmatrix}$$

The indicator weights reported for the comprehensive similarity calculation are W = (0.15, 0.20, 0.10, 0.20, 0.20, 0.15), in the order Euclidean distance, grey relational degree, cosine similarity, DTW, Mahalanobis distance and Spearman correlation. Table 2 summarizes these weights and the role of each indicator.

The consistency of the pairwise comparison matrix is assessed using equations (14) and (15), with a random consistency index of 1.24 for a matrix of order six. The consistency ratio is below 0.1.

Historical days with a comprehensive similarity score above 0.80 are selected for each target date. Each similarity indicator is normalized before the scores are combined using the weights in Table 2. Figure 2a summarizes the mean comprehensive similarity of the selected days, and Fig. 2b shows their score distributions. These selected samples form the inputs to the subsequent decomposition and forecasting stages. The present experiments do not separately compare alternative day-selection or weighting strategies, so their individual contributions to forecast accuracy cannot be isolated from the reported results.

Forecasting model configuration

To ensure the stability and prediction accuracy of DL models in PV power forecasting, systematic debugging and optimization are performed on the model parameters in this study. Considering the nonlinear and multi-scale characteristics of PV power series, long short-term memory (LSTM), gated recurrent units (GRUs) and Transformers are selected as the basic time series forecasting models. Furthermore, combined with the feature extraction capability of the CNN, hybrid models, including CNN-LSTM, CNN-GRU and CNN-Transformer, are constructed to enhance the capture of complex temporal features. The key hyperparameter settings of each model are presented below.

LSTM and GRU models

The core hyperparameters of the LSTM and GRU models are kept consistent to ensure fair comparison. The input sequence length is set to 96, meaning that historical data from the past 96 time steps are adopted for model construction; the prediction sequence length is also defined as 96, corresponding to the power forecasting task for the next 96 time steps. During the training process, the batch size is set to 32, with an initial learning rate of 0.0001. An early stopping strategy is applied to prevent overfitting, and the training is terminated if the validation loss fails to decline for 50 consecutive epochs.

Transformer model

The key configurations of the Transformer model are described below. This study adopts the multivariate-to-univariate (MS) forecasting paradigm, which takes multisource meteorological and power features as model inputs and outputs the target PV power sequence. The rectified linear unit (ReLU) is selected as the activation function. The number of multihead attention heads is set to 8, with 2 encoder layers and 1 decoder layer, and the hidden layer dimension is 512 to balance model capacity and computational efficiency. To suppress overfitting, the dropout rate is set to 0.05. The label sequence length is 48, whereas the input sequence length and prediction horizon are both set to 96. Other training hyperparameters, including the learning rate, batch size and early stopping patience, remain consistent with those of the LSTM and GRU models.

Convolutional hybrid models

In the CNN-LSTM, CNN-GRU, and CNN-Transformer hybrid models, a 1D convolutional layer is introduced for local feature extraction. The convolution kernel size is set to 3 with a stride of 1 and a padding of 1. This configuration ensures that the length of the feature sequence remains unchanged before and after the convolution operation, effectively enhancing local temporal features. The sequence is then fed into the corresponding base model for subsequent time series modeling.

All models use mean squared error as the training loss, with network weights updated by backpropagation. Input features are standardized using Z scores to give zero mean and unit standard deviation.

The model configurations used in the reported comparisons are summarized in Supplementary Table S2. These settings define the configurations evaluated in this study; the comparison does not establish globally optimal hyperparameters.

Manuscript preparation

Codex (OpenAI) assisted with journal formatting, language editing and consistency checks of the manuscript against its reported numerical tables. It was not used in this revision to generate experimental observations or run the forecasting models. The authors are responsible for the final manuscript and its scientific content.

Results

Comparison of decomposition strategies

The power components decomposed by single VMD, single CEEMDAN, and Bayesian-optimized hybrid decomposition are separately imported into the CNN-Transformer hybrid forecasting model. According to the overall forecasting results of historical target days, different decomposition strategies substantially affect model prediction performance, as illustrated by the forecasting curves of the three decomposition schemes in Fig. 3.

With single VMD, CNN–Transformer gives an MAE of 0.5309 and an RMSE of 0.7282. With single CEEMDAN, the corresponding values are 0.6835 and 0.9081 (Table 3). Both configurations have larger errors than the hybrid decomposition configuration.

The Bayesian optimized hybrid strategy first decomposes the power series using VMD and then refines high-complexity components using CEEMDAN. As reported in Table 3, its MAE is 0.4761, RMSE is 0.6582, RSE is 0.1111 and R^2 is 0.8888. Relative to single VMD, MAE and RMSE decrease by 10.3% and 9.6%, respectively. Relative to single CEEMDAN, they decrease by 30.3% and 27.5%. These percentages are calculated from the four-decimal values in Table 3.

For the selected dataset and target period, the results in Table 3 show lower forecast errors for CNN–Transformer with the hybrid decomposition input than with either single decomposition method. This comparison evaluates the complete decomposition configurations and does not separately isolate the contribution of every processing step.

After analyzing the PV power of similar days screened for the final target date (30 June 2019) under different decomposition methods, the complex fluctuation characteristics of the power series present differentiated modal distributions under various decomposition strategies. Taking the historical power data of this day as an example, single VMD divides the original sequence into six modal components and one residual term. Among them, the first three components mainly reflect the intraday smooth trend and periodic fluctuations, whereas the last three contain substantial high-frequency noise, which may be more difficult for the subsequent forecasting model to learn, as shown in Supplementary Fig. S4 for the VMD of PV power.

The single CEEMDAN decomposition yields eight IMF components and one residual term. Although more modes are obtained, obvious information overlap exists among partial high-frequency components with highly similar fluctuation characteristics, which may add redundant information for the subsequent model, as illustrated in Supplementary Fig. S5 for the CEEMDAN decomposition of PV power.

In comparison, the Bayesian-optimized hybrid decomposition strategy first decomposes the original power series into four basic modes via VMD. On the basis of the complexity evaluation utilizing MPE, the high-complexity modes are further secondarily decomposed and reconstructed via CEEMDAN,

ultimately acquiring six components, as presented in Supplementary Fig. S6 for the hybrid decomposition results.

The decomposed series display variation at different apparent timescales (Supplementary Figs. S4–S6). Smooth components may represent intraday trends, while more rapidly varying components may contain weather-related fluctuations or noise. These interpretations are qualitative because the experiments do not separately identify the physical origin of each component. The forecasting comparisons evaluate the complete decomposed inputs and therefore do not isolate the contributions of individual frequency components or architectural modules.

Comparison of forecasting models

The six forecasting models are compared using the same hybrid decomposition input (Table 4 and Fig. 4). The LSTM model has an MAE of 0.5762 and an RMSE of 0.8026, while GRU gives 0.5236 and 0.7250. The standalone Transformer has an MAE of 0.5971 and an RMSE of 0.8457. CNN–LSTM reduces these values to 0.5285 and 0.7414. CNN–GRU gives the lowest MAE, 0.4482, with an RMSE of 0.6861. CNN–Transformer gives an MAE of 0.4761 and the lowest RMSE, 0.6582.

CNN–Transformer also has the lowest RSE, 0.1111, and the highest R^2, 0.8888, among the evaluated models. CNN–GRU gives an RSE of 0.1208 and an R^2 of 0.8791. The distinction between the two models indicates that ranking depends on the error criterion: CNN–GRU has smaller average absolute deviations, whereas CNN–Transformer performs better under the squared-error metrics.

Relative to CNN–LSTM, CNN–Transformer reduces MAE by 9.9% and RMSE by 11.2%; R^2 increases from 0.8589 to 0.8888, an absolute increase of 0.0299. Relative to the standalone Transformer, MAE, RMSE and RSE decrease by 20.3%, 22.2% and 39.5%, respectively. These comparisons support combining convolutional local feature extraction with attention-based temporal modelling for this dataset, while the lower MAE of CNN–GRU shows that CNN–Transformer does not dominate every metric.

Discussion

For the evaluated SOLETE data and target week, CNN–Transformer with the hybrid decomposition input has lower forecast errors than its counterparts using VMD or CEEMDAN alone. The corresponding reductions in MAE are 10.3% and 30.3%. These results describe the tested configurations and period.

The combination of convolutional feature extraction and attention is intended to represent local changes and longer temporal dependencies. CNN–Transformer gives the lowest RMSE and RSE and the highest R^2 in Table 4. CNN–GRU gives the lowest MAE. Thus, model choice should reflect the operational error criterion rather than a claim that one architecture is best under every metric. For the CNN–Transformer configuration, the reported MAE, RMSE, RSE and R^2 are 0.4761, 0.6582, 0.1111 and 0.8888, respectively.

The evaluation is limited to the SOLETE dataset from Denmark and the target week of 24–30 June 2019. The reported metrics are point estimates: repeated-run variability, confidence intervals and statistical significance tests are not reported. The differences therefore describe this evaluation and do not establish robustness across random initializations, additional sites, seasons or extreme-weather regimes. The experiments do not separately isolate the effects of similar-day weighting, Bayesian optimization, the entropy threshold or each network module; targeted ablation studies are needed to assess those contributions. End-to-end runtime, memory use and deployment latency have not been benchmarked here. Operational use also requires time-respecting validation of preprocessing and confirmation that all target-day input variables are available when the forecast is issued. These issues

limit conclusions about generalization, the mechanisms underlying the observed differences and real-time deployment.

Future work can investigate weather-dependent similarity weights, robustness during rapidly changing cloud conditions, transfer learning across sites and seasons, and lighter decomposition procedures. These extensions would test the scope of the observed forecasting improvements and the practicality of the method for routine photovoltaic operation.

Data availability

The SOLETE dataset analysed during the current study is publicly available in the DTU Data repository at https://doi.org/10.11583/DTU.17040767.

Code availability

The preprocessing code and model implementation are available from the corresponding author upon reasonable request.

References

1. Lan, X. et al. Superionic composite electrolytes with continuously perpendicular-aligned pathways for pressure-less all-solid-state lithium batteries. *Nat. Nanotechnol.* **21,** 388–396 (2026). https://doi.org/10.1038/s41565-025-02106-9

2. Li, Z. et al. Boosting mechanical durability under high humidity by bioinspired multisite polymer for high-efficiency flexible perovskite solar cells. *Nat. Commun.* **16,** 1771 (2025). https://doi.org/10.1038/s41467-025-57102-3

3. Li, Z. et al. Hyperbranched polymer functionalized flexible perovskite solar cells with mechanical robustness and reduced lead leakage. *Nat. Commun.* **14,** 6451 (2023). https://doi.org/10.1038/s41467-023-41931-1

4. Zhang, Z. et al. Worldwide rooftop photovoltaic electricity generation may mitigate global warming. *Nat. Clim. Change* **15,** 393–402 (2025). https://doi.org/10.1038/s41558-025-02276-3

5. Estevez, P. G., Marchi, P., Galarza, C. & Elizondo, M. Non-stationary power system forced oscillation analysis using synchrosqueezing transform. *IEEE Trans. Power Syst.* **36,** 1583–1593 (2021). https://doi.org/10.1109/TPWRS.2020.3015145

6. Wen, X. et al. Nonstandard power grid frequency statistics across continents. *Sci. Rep.* **15,** 38470 (2025). https://doi.org/10.1038/s41598-025-25334-4

7. Khetarpal, P., Nagpal, N., Alhelou, H. H., Siano, P. & Al-Numay, M. Noisy and non-stationary power quality disturbance classification based on adaptive segmentation empirical wavelet transform and support vector machine. *Comput. Electr. Eng.* **118,** 109346 (2024). https://doi.org/10.1016/j.compeleceng.2024.109346

8. Mayer, M. J. & Gróf, G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl. Energy* **283,** 116239 (2021). https://doi.org/10.1016/j.apenergy.2020.116239

9. Santos, L. d. O., AlSkaif, T., Barroso, G. C. & Carvalho, P. C. M. d. Photovoltaic power estimation and forecast models integrating physics and machine learning: a review on hybrid techniques. *Sol. Energy* **284,** 113044 (2024). https://doi.org/10.1016/j.solener.2024.113044

10. Kotamraju, S., Sukeerthi, M. & Puthanveettil, S. E. Modeling of InGaP/InGaAs-GaAsP/Ge multiple quantum well solar cell to improve efficiency for space applications. *Sol. Energy* **186,** 328–334 (2019). https://doi.org/10.1016/j.solener.2019.05.015

11. Iversen, E. B., Morales, J. M., Møller, J. K. & Madsen, H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *Int. J. Forecast.* **32,** 981–990 (2016). https://doi.org/10.1016/j.ijforecast.2015.03.001

12. Fentis, A., Bahatti, L., Tabaa, M. & Mestari, M. Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations. *Int. J. Energy Environ. Eng.* **10,** 189–206 (2019). https://doi.org/10.1007/s40095-018-0293-5

13. Voyant, C. et al. Machine learning methods for solar radiation forecasting: a review. *Renew. Energy* **105,** 569–582 (2017). https://doi.org/10.1016/j.renene.2016.12.095

14. Ferkous, K., Guermoui, M., Menakh, S., Bellaour, A. & Boulmaiz, T. A novel learning approach for short-term photovoltaic power forecasting - a review and case studies. *Eng. Appl. Artif. Intell.* **133,** 108502 (2024). https://doi.org/10.1016/j.engappai.2024.108502

15. Mellit, A. & Kalogirou, S. Forecasting of photovoltaic output power using machine learning and deep learning algorithms. In *Handbook of Artificial Intelligence Techniques in Photovoltaic Systems* 113–148 (Elsevier, 2022). https://doi.org/10.1016/B978-0-12-820641-6.00004-1

16. Dhaked, D. K. et al. Exploring deep learning methods for solar photovoltaic power output forecasting: a review. *Renew. Energy Focus* **53,** 100682 (2025). https://doi.org/10.1016/j.ref.2025.100682

17. Khouili, O. et al. Evaluating the impact of deep learning approaches on solar and photovoltaic power forecasting: a systematic review. *Energy Strategy Rev.* **59,** 101735 (2025). https://doi.org/10.1016/j.esr.2025.101735

18. Zhan, Y., Wang, X., Xu, Y. & Li, W. A hybrid TCN-LSTM-attention framework for multi-scenario short-term photovoltaic power forecasting incorporating physics-informed neural network strategy. *Energy* **344,** 139968 (2026). https://doi.org/10.1016/j.energy.2026.139968

19. Yang, X., Wang, S., Peng, Y., Chen, J. & Meng, L. Short-term photovoltaic power prediction with similar-day integrated by BP-adaboost based on the grey-Markov model. *Electr. Power Syst. Res.* **215,** 108966 (2023). https://doi.org/10.1016/j.epsr.2022.108966

20. Huang, X. et al. A short-term power prediction method for photovoltaics based on similar day clustering and spatio-temporal feature extraction. *Electronics* **13,** 3536 (2024). https://doi.org/10.3390/electronics13173536

21. Xiang, C., Liu, X., Liu, W. & Yang, T. A cascaded data-driven approach for photovoltaic power output forecasting. *Mathematics* **13,** 2728 (2025). https://doi.org/10.3390/math13172728

22. Dai, G., Luo, S., Chen, H. & Ji, Y. Efficient method for photovoltaic power generation forecasting based on state space modeling and BiTCN. *Sensors* **24,** 6590 (2024). https://doi.org/10.3390/s24206590

23. Abed, J. N., Abdoos, A. A. & Ebadi, A. A new hybrid intelligent method for short-term photovoltaic power forecasting based on combination of ELM, VMD and PCA. *Electr. Power Syst. Res.* **250,** 112104 (2026). https://doi.org/10.1016/j.epsr.2025.112104

24. Li, Y., Ding, Z., Yu, Y. & Liu, Y. Hybrid energy storage power allocation strategy based on parameter-optimized VMD algorithm for marine micro gas turbine power system. *J. Energy Storage* **73,** 109189 (2023). https://doi.org/10.1016/j.est.2023.109189

25. Liu, L., Zhang, J. & Xue, S. Photovoltaic power forecasting: using wavelet threshold denoising combined with VMD. *Renew. Energy* **249,** 123152 (2025). https://doi.org/10.1016/j.renene.2025.123152

26. Zhai, C. et al. Photovoltaic power forecasting based on VMD-SSA-Transformer: multidimensional analysis of dataset length, weather mutation and forecast accuracy. *Energy* **324,** 135971 (2025). https://doi.org/10.1016/j.energy.2025.135971

27. Zhao, Y. et al. WOA-VMD-SCINet: hybrid model for accurate prediction of ultra-short-term photovoltaic generation power considering seasonal variations. *Energy Rep.* **12,** 3470–3487 (2024). https://doi.org/10.1016/j.egyr.2024.09.025

28. Cui, S., Lyu, S., Ma, Y. & Wang, K. Improved informer PV power short-term prediction model based on weather typing and AHA-VMD-MPE. *Energy* **307,** 132766 (2024). https://doi.org/10.1016/j.energy.2024.132766

29. Zhang, Z. et al. Short-term wind power prediction based on GWO-VMD-FE and TCN-BiGRU. *Electr. Power Syst. Res.* **253,** 112568 (2026). https://doi.org/10.1016/j.epsr.2025.112568

30. Li, Y., Zhang, Y. & Yin, M. Physics-informed mamba network for ultra-short-term photovoltaic power forecasting: integrating WGAN-GP augmentation and CEEMDAN-SST decomposition. *Renew. Energy* **257,** 124851 (2026). https://doi.org/10.1016/j.renene.2025.124851

31. Li, Q., Wang, G., Wu, X., Gao, Z. & Dan, B. Arctic short-term wind speed forecasting based on CNN-LSTM model with CEEMDAN. *Energy* **299,** 131448 (2024). https://doi.org/10.1016/j.energy.2024.131448

32. Karijadi, I., Chou, S.-Y. & Dewabharata, A. Wind power forecasting based on hybrid CEEMDAN-EWT deep learning method. *Renew. Energy* **218,** 119357 (2023). https://doi.org/10.1016/j.renene.2023.119357

33. Li, K., Huang, W., Hu, G. & Li, J. Ultra-short term power load forecasting based on CEEMDAN-SE and LSTM neural network. *Energy Build.* **279,** 112666 (2023). https://doi.org/10.1016/j.enbuild.2022.112666

34. Liang, J. et al. Short-term photovoltaic power prediction based on CEEMDAN-PE and BiLSTM neural network. *Electr. Power Syst. Res.* **246,** 111706 (2025). https://doi.org/10.1016/j.epsr.2025.111706

35. Wang, K., Qi, X. & Liu, H. Photovoltaic power forecasting based LSTM-convolutional network. *Energy* **189,** 116225 (2019). https://doi.org/10.1016/j.energy.2019.116225

36. Qu, J., Qian, Z. & Pei, Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* **232,** 120996 (2021). https://doi.org/10.1016/j.energy.2021.120996

37. Oumiguil, L. & Nejmi, A. Short-term PV plant power forecasting model using CNN-GRU with xgboost algorithm. *Sci. Afr.* **31,** e03184 (2026). https://doi.org/10.1016/j.sciaf.2026.e03184

38. Liu, M., Rao, S., Huang, M. & Deng, S. Short-term photovoltaic power forecasting based on improved transformer with feature enhancement. *Sustain. Energy Grids Netw.* **43,** 101759 (2025). https://doi.org/10.1016/j.segan.2025.101759

39. Liu, X., Wang, Z., Wang, Y., Liu, S. & Yu, L. A hybrid ultra-short-term photovoltaic power prediction framework integrating ant colony optimization for clustering with bi-GRU. *Int. J. Electr. Power Energy Syst.* **170,** 110905 (2025). https://doi.org/10.1016/j.ijepes.2025.110905

40. Fan, S., Geng, H., Zhang, H., Yang, J. & Hiroichi, K. Photovoltaic power forecasting model employing epoch-dependent adaptive loss weighting and data assimilation. *Sol. Energy* **290,** 113351 (2025). https://doi.org/10.1016/j.solener.2025.113351

41. Pombo, D. V. et al. Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting. *Sustain. Energy Grids Netw.* **32,** 100943 (2022). https://doi.org/10.1016/j.segan.2022.100943

42. Harary, M. Efficient algorithms for the sensitivities of the Pearson correlation coefficient and its statistical significance to online data. Preprint at https://arxiv.org/abs/2405.14686 (2024).

43. Salomon, V. A. P. & Gomes, L. F. A. M. Consistency improvement in the analytic hierarchy process. *Mathematics* **12,** 828 (2024). https://doi.org/10.3390/math12060828

44. Schmidt, K., Aumann, I., Hollander, I., Damm, K. & von der Schulenburg, J. M. G. Applying the analytic hierarchy process in healthcare research: a systematic literature review and evaluation of reporting. *BMC Med. Inform. Decis. Mak.* **15,** 112 (2015). https://doi.org/10.1186/s12911-015-0234-7

45. Cooper, O. The Magic of the analytic hierarchy process (AHP). *Int. J. Anal. Hierarchy Process* **9,** (2017). https://doi.org/10.13033/ijahp.v9i3.519

46. Dragomiretskiy, K. & Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **62,** 531–544 (2014). https://doi.org/10.1109/TSP.2013.2288675

47. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems* **25** (Curran Associates, 2012). https://proceedings.neurips.cc/paper_files/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html

48. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & de Freitas, N. Taking the human out of the loop: a review of Bayesian optimization. *Proc. IEEE* **104,** 148–175 (2016). https://doi.org/10.1109/jproc.2015.2494218

49. Wang, Z., Hutter, F., Zoghi, M., Matheson, D. & de Freitas, N. Bayesian optimization in a billion dimensions via random embeddings. *J. Artif. Intell. Res.* **55,** 361–387 (2016). https://doi.org/10.1613/jair.4806

50. Eriksson, D., Pearce, M., Gardner, J. R., Turner, R. D. & Poloczek, M. Scalable global optimization via local Bayesian optimization. In *Advances in Neural Information Processing Systems* **32** (Curran Associates, 2019). https://proceedings.neurips.cc/paper_files/paper/2019/hash/6c990b7aca7bc7058f5e98ea909e924b-Abstract.html

51. Zhou, Z., Chen, W. & Yang, C. Adaptive range selection for parameter optimization of VMD algorithm in rolling bearing fault diagnosis under strong background noise. *J. Mech. Sci. Technol.* **37,** 5759–5773 (2023). https://doi.org/10.1007/s12206-023-1015-3

52. Wang, X., Si, S., Wei, Y. & Li, Y. The optimized multi-scale permutation entropy and its application in compound fault diagnosis of rotating machinery. *Entropy* **21,** 170 (2019). https://doi.org/10.3390/e21020170

53. Dávalos, A., Jabloun, M., Ravier, P. & Buttelli, O. On the statistical properties of multiscale permutation entropy: characterization of the estimator's variance. *Entropy* **21,** 450 (2019). https://doi.org/10.3390/e21050450

54. Fan, C.-L., Jin, N.-D., Chen, X.-T. & Gao, Z.-K. Multi-scale permutation entropy: a complexity measure for discriminating two-phase flow dynamics. *Chin. Phys. Lett.* **30,** 090501 (2013). https://doi.org/10.1088/0256-307x/30/9/090501

55. Torres, M. E., Colominas, M. A., Schlotthauer, G. & Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing* 4144–4147 (IEEE, 2011). https://doi.org/10.1109/ICASSP.2011.5947265

56. Pombo, D. V., Gehrke, O. & Bindner, H. W. SOLETE, a 15-month long holistic dataset including: meteorology, co-located wind and solar PV power from Denmark with various resolutions. *Data Brief* **42,** 108046 (2022). https://doi.org/10.1016/j.dib.2022.108046

Acknowledgements

We thank the SOLETE data contributors for making the photovoltaic and meteorological observations available, and our laboratory colleagues for discussions.

Funding

This research was funded by the General Project of the National Social Science Fund of China, "An Empirical Examination and Optimization of the Deterrent System for Aggravated Damages in the Context of Strict Intellectual Property Protection" (grant no. 25BFX180). Article processing charge funding was provided by Yuan Xue.

Author contributions

Chenyu Qi: Conceptualization, methodology, data curation, formal analysis, visualization, writing—original draft, and writing—review and editing. Hua He: Data curation, investigation, resources, and writing—review and editing. Yuan Xue: Formal analysis, visualization, and writing—review and editing. All authors have read and approved the manuscript.

Additional information

Competing interests

The authors declare no competing interests.

Supplementary information

Supplementary Information contains the Bayesian optimization workflow, feature correlation plots, decomposition examples, the AHP scale and model parameter configuration.

Correspondence

Correspondence and requests for materials should be addressed to Yuan Xue (xueyuan95@shu.edu.cn).

Figure legends

Figure 1. Forecasting framework. Feature screening, comprehensive similar-day selection, selective hybrid signal decomposition and forecasting model comparison. VMD, variational mode decomposition; MPE, multiscale permutation entropy; CEEMDAN, complete ensemble empirical mode decomposition with adaptive noise; CNN, convolutional neural network; LSTM, long short-term memory; GRU, gated recurrent unit.

Figure 2. Selection of historical similar days. (a) Mean comprehensive similarity for historical days selected for each target date from 24 to 30 June 2019. (b) Distribution of comprehensive similarity scores for the selected historical days. The selection threshold is 0.80.

Figure 3. Forecasts obtained with three decomposition strategies. Observed photovoltaic power and CNN–Transformer forecasts using single VMD, single CEEMDAN and Bayesian optimized VMD–MPE–CEEMDAN input over the target period of 24–30 June 2019.

Figure 4. Comparison of forecasting models with hybrid decomposition input. (a) Observed and predicted photovoltaic power for LSTM, GRU, Transformer, CNN–LSTM, CNN–GRU and CNN–Transformer. (b) Daily observed and CNN–Transformer predicted profiles for the target dates of 24–30 June 2019. Panel (b) overlays the separate daily profiles to show differences in intraday shape.

Table 1. Definitions of candidate forecasting features.

Feature Parameter	Definition	Unit
TEMPMODULE_RP	Module temperature corrected by relative humidity	°C
WIND_SPEED	Ambient wind speed at the current moment	m/s
WIND_DIR	Ambient wind direction at the current moment	deg
AZIMUTH	Solar azimuth angle at the current moment	deg
PRESSURE	Ambient atmospheric pressure	mbar
TEMPERATURE	Ambient temperature	°C
STDPREVH	Standard deviation of photovoltaic power in the previous hour	kW
HUMIDITY	Ambient relative humidity	%
ELEVATION	Solar altitude angle at the current moment	deg
TEMPMODULE	Photovoltaic module temperature	°C
TEMPCELL	Solar cell temperature	°C
GHI	Global horizontal irradiance	kW/m²
MEANPREVH	Average photovoltaic power in the previous hour	kW
POA_IRR	Plane-of-array irradiance	kW/m²

Table 2. Weights of the six similarity indicators.

Similarity Indicator	Weight Value	Indicator Description
Euclidean Distance	0.15	Reflects the absolute numerical difference; serves as a fundamental indicator yet susceptible to extreme values.
Grey Correlation Degree	0.20	Highlights overall trend similarity and exhibits good adaptability to nonlinear and small-sample data, with high importance in photovoltaic power forecasting.
Cosine Similarity	0.10	Focuses on directional consistency and is sensitive to morphological similarity, acting as a supplement for trend analysis.
DTW	0.20	Eliminates time misalignment problems caused by peak and valley offset of meteorological data, and improves the robustness of similarity evaluation.
Mahalanobis Distance	0.20	Eliminates the interference of dimensional differences and feature correlation, and optimizes distance measurement in the feature space.
Spearman Correlation Coefficient	0.15	Captures nonlinear monotonic trends, which is consistent with the monotonic correlation between photovoltaic output and typical meteorological features such as irradiance.

Table 3. Forecasting metrics for CNN–Transformer with three decomposition strategies.

Decomposition Method	MAE	RMSE	RSE	R^2
VMD	0.5309	0.7282	0.1361	0.8638
CEEMDAN	0.6835	0.9081	0.2116	0.7883
Bayesian optimized hybrid	0.4761	0.6582	0.1111	0.8888

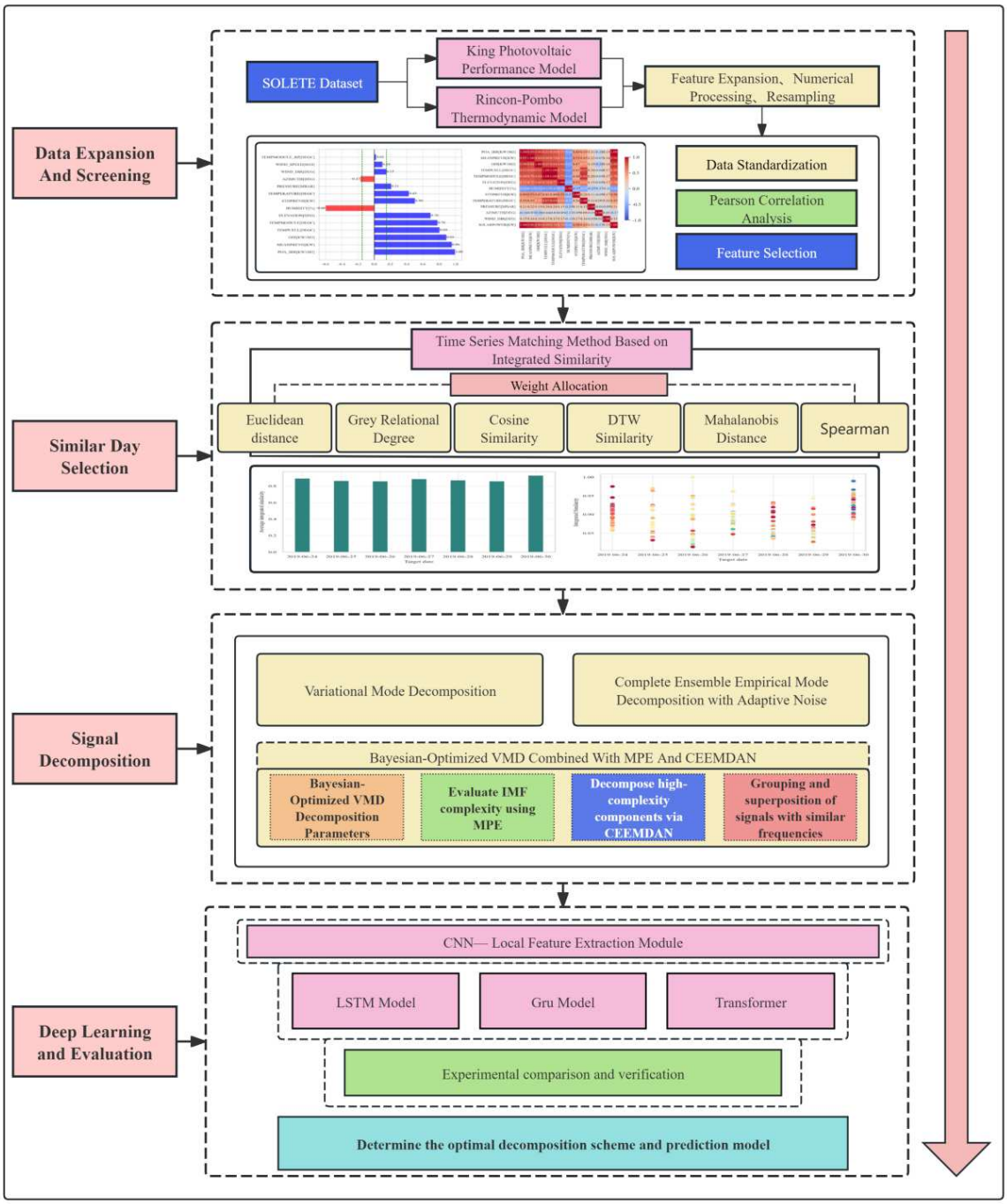
MAE, mean absolute error; RMSE, root mean square error; RSE, relative squared error; R^2, coefficient of determination.

Table 4. Forecasting metrics for six models with hybrid decomposition input.

Model	MAE	RMSE	RSE	R^2
LSTM	0.5762	0.8026	0.1653	0.8346
GRU	0.5236	0.7250	0.1349	0.8650
Transformer	0.5971	0.8457	0.1835	0.8164
CNN–LSTM	0.5285	0.7414	0.1410	0.8589
CNN–GRU	0.4482	0.6861	0.1208	0.8791
CNN–Transformer	0.4761	0.6582	0.1111	0.8888

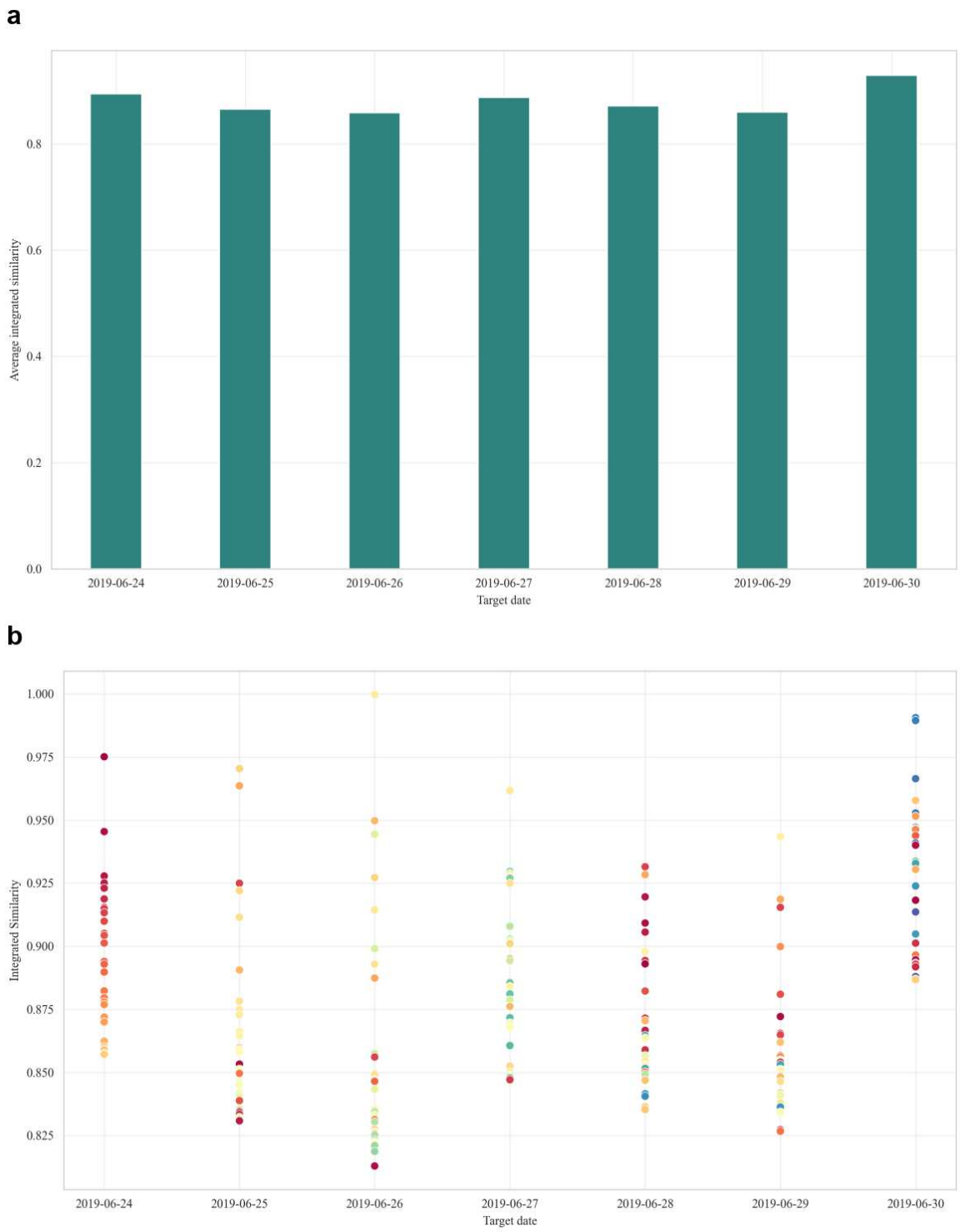
MAE, mean absolute error; RMSE, root mean square error; RSE, relative squared error; R^2, coefficient of determination.

Figure 1. Forecasting framework.



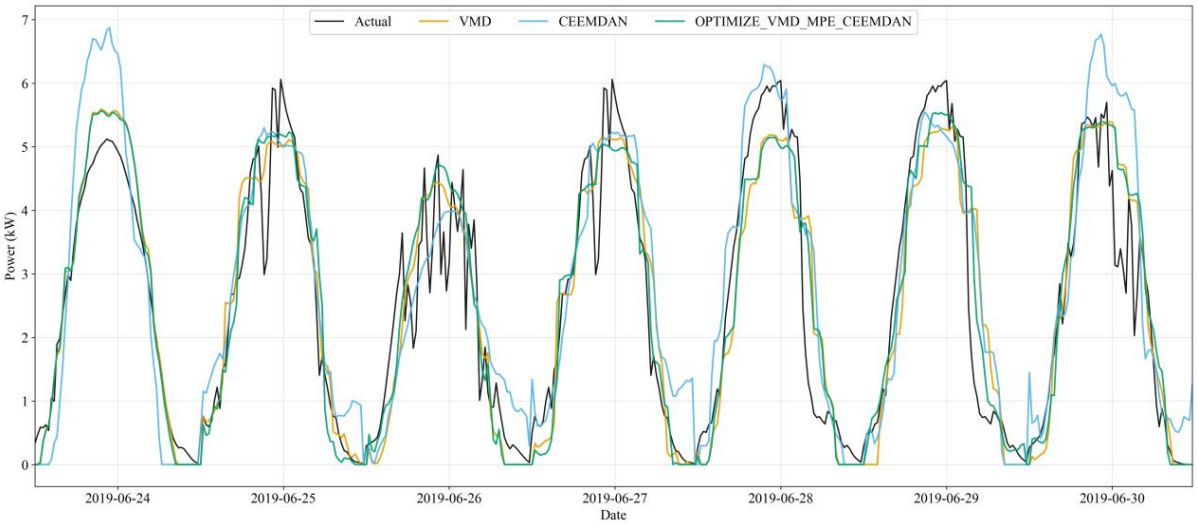
Feature screening, comprehensive similar-day selection, selective hybrid signal decomposition and forecasting model comparison. VMD, variational mode decomposition; MPE, multiscale permutation entropy; CEEMDAN, complete ensemble empirical mode decomposition with adaptive noise; CNN, convolutional neural network; LSTM, long short-term memory; GRU, gated recurrent unit.

Figure 2. Selection of historical similar days.



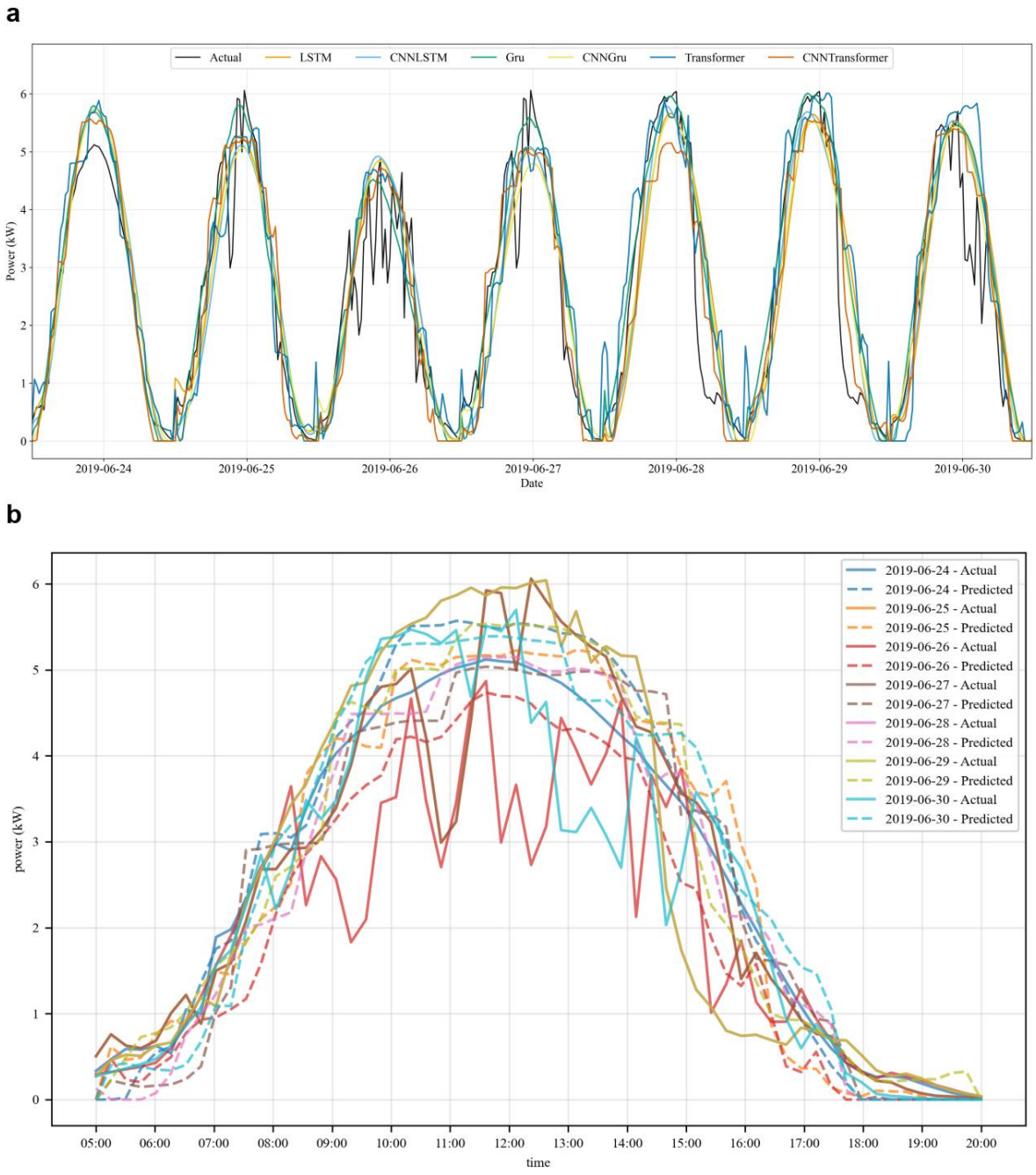
(a) Mean comprehensive similarity for historical days selected for each target date from 24 to 30 June 2019. (b) Distribution of comprehensive similarity scores for the selected historical days. The selection threshold is 0.80.

Figure 3. Forecasts obtained with three decomposition strategies.



Observed photovoltaic power and CNN–Transformer forecasts using single VMD, single CEEMDAN and Bayesian optimized VMD–MPE–CEEMDAN input over the target period of 24–30 June 2019.

Figure 4. Comparison of forecasting models with hybrid decomposition input.



(a) Observed and predicted photovoltaic power for LSTM, GRU, Transformer, CNN–LSTM, CNN–GRU and CNN–Transformer. (b) Daily observed and CNN–Transformer predicted profiles for the target dates of 24–30 June 2019. Panel (b) overlays the separate daily profiles to show differences in intraday shape.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx