EAI Endorsed Transactions
on Energy Web
Research Article EAI.EU

# Deep Clustering and Adaptive Mode Decomposition of Industrial Loads for Temporal Complexity

Shun Yan[1,2,*], Li Lv[1,2], Zengyao Tian[1,2], Qun Wang[1,2], Yanqing Zhao[1,2]

[1]Shenyang Institute of Computing Technology, Chinese Academy of Sciences, Shenyang 110168, China
[2]University of Chinese Academy of Sciences, Beijing 101408, China

## Abstract

Accurate extraction of underlying features from electrical loads is crucial for intelligent microgrid management.To address the heterogeneous characteristics and microscopic nonstationary properties of industrial loads, a shunt-cooperative feature extraction framework relying on MSE is proposed in this work. Different from traditional methods that use a single feature extraction or clustering model to process the entire sequence, MSE is introduced to quantify the dynamic complexity of load sequences, and effective classification is performed for loads with strong and weak time-series characteristics. For strong time-series loads, a multi-view Variational AutoEncoder (VAE) deep clustering model with adaptive dual-distance metrics fused with Euclidean distance and LB_Keogh distance is constructed to accurately extract macroscopic typical power consumption patterns. For weak time-series loads, a "coarse screening-fine classification" hybrid decomposition architecture based on CEEMDAN-NGO-VMD is proposed, where the complexity index (LZC) is introduced to adaptively strip low-frequency trends and Northern Goshawk Optimization (NGO) is used to optimize VMD and suppress high-frequency mode aliasing. Simulation results show that, compared with the CEEMDAN-only method and the NGO-VMD-only algorithm, the proposed method achieves accurate extraction of multi-scale time-frequency features and effectively overcomes the mode aliasing problem of traditional decomposition methods in terms of clustering performance. Finally, complex loads are reconstructed into multi-modal feature components with clear physical meanings, providing a reliable feature basis for subsequent identification of multi-modal power consumption behaviors in industrial microgrids.

**Keywords:** Industrial Microgrid, Multiscale Sample Entropy, Multi-View Variational Autoencoder, Deep Clustering, CEEMDAN-NGO-VMD, Temporal Complexity

Received on 27 June 2026, accepted on 21 September 2026, published on 05 November 2026

Copyright © 2026 Shun Yan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/ew.16579

## 1. Introduction

In industrial microgrids, deep coupling of high-proportion distributed new energy and energy storage is observed, where advanced deep learning frameworks have demonstrated superior performance in pattern recognition and data analysis [1]. Industrial power consumption behavior is presented as a nonlinear superposition of macroscopic production rhythms and microscopic random fluctuations, showing extremely high complexity. Effective analysis of big data is considered significant for the optimization of power resource allocation. Therefore, an in-depth analysis of load characteristics is required. Accurate characterization of load evolution rules is realized through effective feature extraction and information mining, which provides a reliable feature basis for subsequent identification of multi-modal power consumption behaviors. Recent studies on high-energy-consuming industrial plants have shown that high-dimensional daily load sequences contain complex nonlinear characteristics, making effective

*Corresponding author. Email: 18242805177@163.com

feature extraction and clustering essential for accurately characterizing industrial electricity consumption patterns [2]. The integration of Artificial Intelligence (AI) and Machine Learning (ML) paradigms, especially deep learning models, has emerged as a pivotal solution for enhancing the resilience and predictive analytics of modern power systems, and deep learning approaches have achieved remarkable performance in load forecasting [3]. Concurrently, to accurately capture the microscopic dynamics inherent in non-stationary power consumption signals, adaptive multi-resolution dynamic mode decomposition frameworks have been introduced to improve feature extraction precision and suppress mode aliasing in complex industrial load profiles [4].

Load feature extraction is employed to identify and quantify key information reflecting users' power consumption patterns, equipment operating status, and demand response potential from massive power consumption data. These features are applied as the basis for load clustering analysis, demand response potential assessment, non-intrusive load decomposition, and other applications. Especially in the context of high-proportion renewable energy connected to the grid, accurate mining of the flexibility potential of industrial users is considered crucial for maintaining the stability of microgrids. Early studies are mainly focused on clustering based on traditional statistical features. Scalar features of load curves, such as daily load factor, peak-to-valley ratio, and extreme loads, are calculated as static indices, upon which conventional algorithms like K-Means [5] and hierarchical clustering [6] are adopted for user grouping. For example, early demand response projects at home and abroad are often used to identify “high-potential interruptible users” based on such clustering results. However, time-series morphological information of loads is lost in such methods, and the ability to distinguish complex nonlinear patterns is limited. In Reference [7], an Adaptive S Transform (AST) method is proposed. The standard deviation of the window function is directly controlled to adjust time-frequency resolution. The Gaussian window is matched with the principal value interval of the power quality signal spectrum to optimize time-frequency resolution, and more accurate feature vectors are extracted. Reference proposes a clustering method based on load time series [8]. Correlation between multiple electricity customers is effectively captured and representative load patterns are extracted. Combined with an alternating optimization algorithm of Lagrange multiplier and eigenvalue decomposition, a globally optimal solution with fast convergence is obtained, and computational efficiency is improved.

Different from clustering analysis, which focuses on the relationship between data samples, modal decomposition [9] is a signal processing method that decomposes complex non-stationary signals into a series of components with different time-frequency characteristics. Entropy-based wavelet packet decomposition has been employed to extract informative frequency-domain features from electrical load series [10]. Meanwhile, singular value decomposition has been utilized to extract dominant characteristics of daily load curves through dimensionality reduction, thereby improving the efficiency and robustness of load clustering [11]. For nonlinear and non-stationary load sequences, a CEEMDAN-based framework combined with entropy-based component screening and secondary VMD has been proposed to further refine complex high-frequency components and enhance the representation of load characteristics [12]. Reference calculates statistical features of each IMF after decomposing power quality disturbance signals with VMD [13]. A two-stage feature selection method based on permutation entropy and Fisher score is introduced. Redundant features are effectively eliminated, and the performance of support vector machine classifiers is improved.

Obvious deficiencies are found in existing methods for processing the above high-dimensional complex loads. Traditional shallow clustering is limited by a single distance metric and ignores multi-scale dynamic correlations, resulting in poor pattern division accuracy [14]. Even with the introduction of improved density-based algorithms [15] or emerging deep clustering frameworks [16], achieving robust characterization in highly non-stationary and high-dimensional environments remains difficult. Furthermore, industrial load sequences are characterized by pronounced nonlinearity, volatility, and randomness, making it difficult for a single decomposition method to sufficiently characterize their complex multi-scale temporal features; combined decomposition strategies involving VMD, CEEMDAN, and complexity evaluation have therefore been introduced for industrial load analysis [17]. Therefore, existing modeling schemes are unable to meet the demand for an in-depth perception of complex industrial power consumption behaviors in industrial microgrids.

To solve the above problems, a shunting collaborative feature extraction framework for complex nonlinear industrial loads in industrial microgrids is proposed in this paper. The problem of high-dimensional feature analysis under the scenario of deep source-load coupling is solved, and the refined perception ability of complex power consumption behaviors is improved.

Different from traditional methods that use a single feature extraction or clustering model to process the entire sequence, MSE is adopted in this framework to quantify the dynamic complexity of loads. Data are divided into strong and weak time-series loads, and feature confusion and pattern overlap caused by hybrid modeling are avoided. A multi-view VAE deep clustering model is constructed. Adaptive dual-distance metrics are integrated, macroscopic magnitude differences and local time-series fluctuations of data are considered, and adaptive extraction of typical power consumption patterns is realized.

Finally, a “coarse screening-fine classification” hybrid decomposition architecture based on CEEMDAN-NGO-VMD is proposed. The Lempel-Ziv Complexity (LZC) index is used to adaptively screen components, and the NGO algorithm is combined to optimize VMD parameters. Different from traditional signal decomposition that is prone to mode aliasing, adaptive decomposition and feature extraction of nonlinear high-frequency components are realized by this method.

The innovations of this paper are listed as follows:

Current feature extraction methods are difficult to take into account the diversity and microscopic non-stationarity of industrial loads. A shunting collaborative analysis framework for industrial loads in microgrids is adopted. MSE is introduced to quantify the dynamic complexity of load sequences, and the definition of strong and weak time-series loads is realized.

For strong time-series loads, a multi-view VAE deep clustering model with adaptive dual-distance metrics fused with Euclidean distance and LB_Keogh distance is constructed to accurately extract macroscopic typical power consumption patterns [18]. For weak time-series loads, a "coarse screening-fine classification" hybrid decomposition architecture based on CEEMDAN-NGO-VMD is proposed. The complexity index (LZC) is introduced to adaptively strip low-frequency trends, and Northern Goshawk Optimization (NGO) is used to optimize VMD and suppress high-frequency mode aliasing [19]. Finally, complex loads are reconstructed into multi-modal feature components with clear physical meanings.

## 2. Load Sequence Complexity Measurement Based on Multi-Scale Sample Entropy

Refined management needs are put forward for massive power users in industrial parks. It is difficult to analyze individual power consumption characteristics only by manual experience. Core load evolution rules are extracted based on historical monitoring data and advanced data mining methods.

Investigation results show that industrial power consumption is essentially different from civil and commercial loads because it is constrained by production scheduling plans and usually has obvious periodic rhythms. Under conventional production processes, power demand is high and stable on working days, and load is reduced on rest days due to decreased production capacity. However, non-periodic random characteristics are shown in load curves of some industrial individuals with continuous operation or those greatly affected by order fluctuations, climate change, and regional policies.

To accurately quantify such heterogeneous load characteristics, MSE is adopted as the evaluation index of load periodicity in this paper. Through coarse-graining processing of load sequences at different time scales, MSE is used to characterize the dynamic complexity of the system from both microscopic fluctuations and macroscopic trends.

### 2.1 Principle of Multi-Scale Sample Entropy

The multiscale sample entropy (MSE) method is proposed to characterize the complexity of time-series signals across different temporal granularities. By introducing a coarse-graining procedure to the standard sample entropy, MSE effectively captures the underlying non-linear dynamics of power consumption. This approach has been widely verified as a robust feature extraction technique for residential and industrial load profile analysis [20]. In the multi-scale coarse-graining process, the sampling interval of industrial load data adopted in this study is 15 minutes. The maximum scale factor is set to 20, corresponding to a time span of minutes. The 5-hour time scale is sufficient to cover and reflect the power consumption rhythm and capacity fluctuation rules of a typical half shift (such as morning shift, middle shift) in industrial production. Accurate capture of dynamic characteristics under real production cycles by MSE is ensured. The calculation steps are given as follows:

(1) For the original industrial load sequence with length N, the scale factor τ is introduced. The original sequence is divided into several non-overlapping windows of length τ. The average value in each window is calculated to construct the coarse-grained sequence x(τ) at scale τ:

$$y_j^{(\tau)} = \frac{1}{\tau}\sum_{i=(j-1)\tau+1}^{j\tau} x_i, \quad 1 \le j \le \frac{N}{\tau} \tag{1}$$

where τ is a continuous positive integer. When τ=1, the coarse-grained sequence is the original sequence.

(2) For the obtained coarse-grained sequence x(τ), the embedding dimension m is set. A series of vectors Xi with dimension m is constructed:

$$Y_m(i) = [y_i, y_{i+1}, \ldots, y_{i+m-1}], 1 \le i \le N_{\tau-m+1} \tag{2}$$

where N(τ) is the total length of the sequence at this scale.

(3) The distance d[$Y_m$(i), $Y_m$(j)] between vectors $Y_m$(i) and $Y_m$(j) is defined as the maximum absolute difference of their corresponding elements:

$$d[Y_m(i), Y_m(j)] = \max_{k=0\ldots m-1}(|y_{i+k} - y_{j+k}|) \tag{3}$$

The similarity tolerance r is specified (generally set to 0.15–0.25 times the standard deviation of the original sequence). The number of vector pairs satisfying d<r is counted, and its ratio to the total number of vectors is calculated as $B_i^m$(r):

$$B_i^m(r) = \frac{1}{N_\tau - m}\text{count}\{d[Y_m(i), Y_m(j)] < r, i \ne j\} \tag{4}$$

The average value over all indices i is further computed:

$$B^m(r) = \frac{1}{N_\tau - m + 1}\sum_{i=1}^{N_\tau - m + 1} B_i^m(r) \tag{5}$$

(4) The embedding dimension is increased to m+1. The above steps are repeated, and the average value Am(r) is obtained:

$$A^m(r) = \frac{1}{N_\tau - m}\sum_{i=1}^{N_\tau - m} A_i^m(r) \tag{6}$$

where Am(r) represents the probability that two sequences match at m+1 points under the tolerance r.

(5) The sample entropy SampEn at scale τ is defined as:

$$SampEn(m,r,\tau) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \quad (7)$$

(6) The scale factor τ is varied from 1 to τmax. Steps (1) to (5) are repeated, and a set of entropy values is acquired.

As indicated by Formula (7), the calculated sample entropy is affected by the selections of parameters m and r. According to existing research, reasonable statistical properties of sample entropy are obtained when m=1 or 2 and r=0.1Std~0.25Std, where Std denotes the standard deviation of the original dataset x(i), i=1,2,…,N. An excessively large m imposes strict requirements on data length, while an overly small m fails to capture the dynamic characteristics of the original time series. The selection of r directly determines entropy stability: an extremely small r results in zero matching pairs and invalid computation; an excessively large r causes loss of detailed signal information. In this paper, parameters are set as m=2, r=0.2Std.

## 2.2 Classification Criteria for Strong/Weak Time-Series Characteristics

Enterprises with fixed production schedules and high automation degrees (such as steel plants and data centers) are defined as strong time-series loads in this paper. Manufacturing enterprises with irregular overtime or severe source-load interaction are categorized as weak time-series loads, whose power consumption is unstable.

The average entropy $\bar{E}$ is adopted as the classification index:

$$\bar{E} = \frac{1}{\tau_{max}} \sum_{\tau=1}^{\tau_{max}} SampEn(m,r,\tau) \quad (8)$$

The smaller the index value, the higher the self-similarity of the load at each scale and the more stable the production law.

The specific judgment index is selected as:

$$\begin{cases} X \in \text{Cstrong(Strong time} - \text{series load)}, & \bar{E} < 1.25 \\ X \in \text{Cweak(Weak time} - \text{series load)}, & \bar{E} \geq 1.25 \end{cases} \quad (9)$$

In this paper, the power consumption of 5 users in the park is statistically analyzed. The average sample entropy and coefficient of variation of daily load sequences within one month are calculated, respectively. The accuracy and sensitivity of the average sample entropy value in the classification of industrial load time-series characteristics are verified.

Figure 1 shows the effectiveness of MSE in measuring the complexity of industrial loads. The trend curves of sample entropy with scale factor for 5 typical industrial users are presented. It can be seen that with the increase of coarse-graining scale, the evolution trajectories of sample entropy of different users show significant differentiation. The sample entropy curves of User 1, User 2, and User 4 are generally at a low level. Their average sample entropy (1.14, 0.65, and 0.88, respectively) is below the classification threshold ($\bar{E}$=1.25). High self-similarity and regularity are maintained by such loads at multiple time scales. On the contrary, the sample entropy curves of User 3 and User 5 are generally at a high level. Strong oscillation or divergence trends are shown with the increase of scale. The average sample entropy of these two users (1.41 and 1.49, respectively) is above the classification threshold (marked by the red line). A large number of nonlinear components and disordered fluctuations are contained in such load sequences.

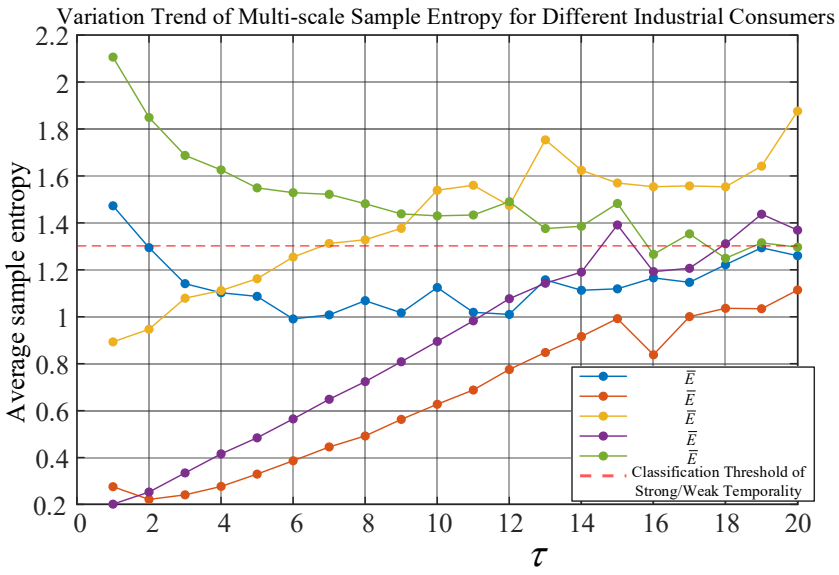



**Figure 1.** Multi-scale sample entropy changes with scale factor

Table 1. Comparison of calculation results of different timing characteristics discriminant indexes

| Industrial User | Average Sample Entropy $\bar{E}$ | Coefficient of Variation $C_v$ |
|---|---|---|
| User 1 | 1.141 | 0.206 |
| User 2 | 0.655 | 0.951 |
| User 3 | 1.408 | 0.064 |
| User 4 | 0.881 | 0.393 |
| User 5 | 1.493 | 0.343 |

It can be seen from Table 1 that although the traditional coefficient of variation can reflect the load fluctuation to a certain extent, limitations are found in distinguishing the types of load fluctuations. The average sample entropy can accurately reflect the nonlinear components in the load sequence. The entropy values of User 2 and User 4 are both lower than 1.25, reflecting their strong production regularity and high time-series characteristics. User 1 is also defined as a strong time-series load. The entropy values of User 3 and User 5 both exceed 1.4 with no obvious production cycle, so they are defined as weak time-series loads. Classification based on $\bar{E}$ can effectively classify low-entropy loads into Category A (strong time-series) for typical pattern extraction, and high-entropy loads into Category B (weak time-series) for independent modeling and analysis of load characteristics.

## 3. Clustering Analysis of Strong Time-Series Loads

After screening the data set by MSE, the power consumption behavior of some users shows strong time-series characteristics due to industrial production scheduling and automated technological processes.

Clustering analysis is carried out for strong time-series loads. Typical daily power consumption patterns representative of the industry are extracted from redundant and high-dimensional historical power sequences, providing benchmark labels for subsequent behavior identification.

In this section, a multi-view deep learning architecture is introduced. The feature decomposition ability of generative models is used to realize efficient and accurate characterization of industrial users' power consumption behaviors.

### 3.1 Efficient Deep Multi-View Clustering of Power Users Based on Variational AutoEncoder

The block diagram of the multi-view VAE is shown in Figure 2, which consists of three modules: a multi-view variational autoencoder network [21], a deep fusion module, and a multi-scale measurement fusion module. The multi-view variational autoencoder network is responsible for data loading and preprocessing. The deep fusion module performs data feature extraction. The multi-scale measurement fusion module realizes clustering of power users. These three modules are adopted to effectively extract valuable information from industrial load data while considering the different characteristics and long-term fluctuations of industrial loads.

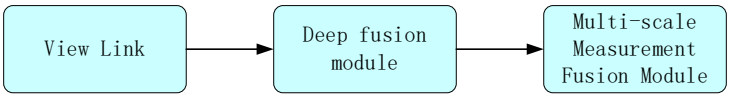



**Figure 2.** Variational autoencoder block diagram

(1) VAE

An autoencoder is a commonly used self-supervised learning model. The overall network framework is shown in Figure 3.

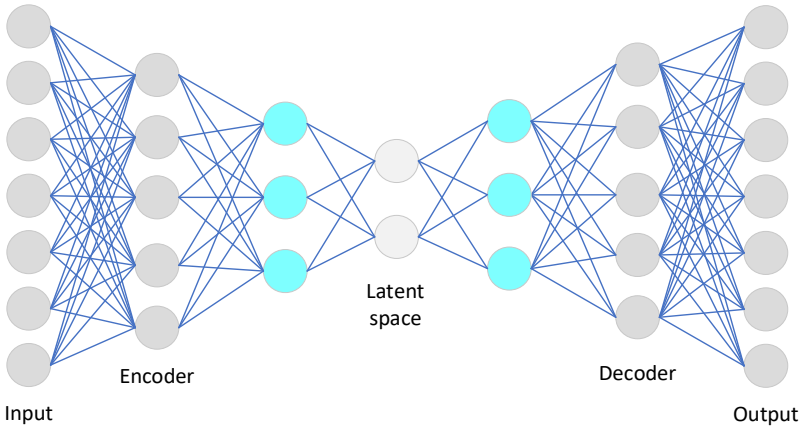



**Figure 3.** Autoencoder network structure

VAE is mainly composed of two modules. An encoder$\left(q_\varphi(z|x)\right)$is adopted to map high-dimensional input data into low-dimensional latent features. A decoder $\left(p_\theta(x|z)\right)$is configured to perform the inverse transformation and reconstruct the original input data from latent features. Better performance is achieved by the variational autoencoder (VAE) compared with conventional autoencoders. Probabilistic distribution models are constructed for the encoder (parameterized by φ) and the decoder (parameterized by θ), denoted as $\left(q_\varphi(z|x)\right)$ and $\left(p_\theta(x|z)\right)$ respectively.

The true posterior distribution $\left(p_\theta(x|z)\right)$ cannot be obtained directly. The approximate posterior $\left(q_\varphi(z|x)\right)$ is employed in VAE for distribution estimation. Sampling from $\left(q_\varphi(z|x)\right)$is used to approximate the sampling process from the true posterior $\left(p_\theta(x|z)\right)$. The Kullback-Leibler (KL) divergence is introduced to quantify the discrepancy between the two distributions for distribution matching. A smaller KL divergence corresponds to higher similarity between two probability distributions. The KL divergence is defined as:

$$D_{KL}(q_\varphi(z \mid x)) \parallel p_\theta(x \mid z)) \tag{10}$$

KL divergence can be minimized by maximizing the evidence lower bound (ELBO). Therefore, the network loss of VAE is expressed as follows:

$$Loss_{VAE} = E_{q_\varphi(z|x)}[log\, p_\theta\,(x|z)] - D_{KL}\left(q_\varphi(z|x)\|p_\theta(x|z)\right) \tag{11}$$

(2) Multi-view VAE Network

The structure of VAE is improved in this paper to capture the unique latent features of different views. The network structure of the multi-view VAE is shown in Figure 4.

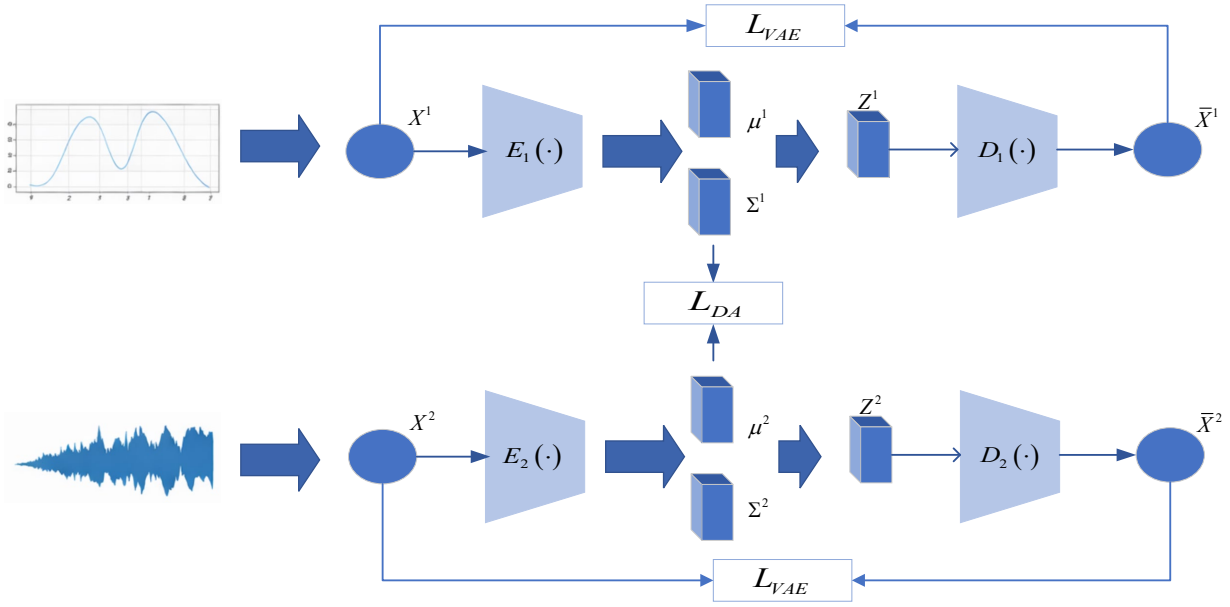



**Figure 4.** Multi-view VAE network structure diagram

Suppose the i-th view is denoted as H(i). A multi-view electrical dataset is composed of V groups of view data {X1, X2, …, XV}, where X(i) stands for the data feature of power load extracted from the i-th view. Based on Formula (11), the loss function of the multi-view autoencoder network is defined as:

$$Loss_{VAE} = \sum_{i=1}^{v} E_{q_\varphi(z^{(i)}|x^{(i)})}\, log[\, p_\theta(x^{(i)} \mid z^{(i)})] - D_{KL}(q_\varphi(z^{(i)} \mid x^{(i)}) \parallel p_\theta(x^{(i)} \mid z^{(i)})) \tag{12}$$

To ensure that the multi-view learning process can capture consistent and complementary information in multi-view data, the Wasserstein distance is adopted to align the Gaussian distribution of latent variable Z among views. Thus, the distribution alignment loss can be defined as:

$$Loss_{DA} = ((\parallel u^{(i)} - u^{(j)} \parallel_F^2) + (\parallel \Sigma_{(i)}^{\frac{1}{2}} - \Sigma_{(j)}^{\frac{1}{2}} \parallel_F^2))^{\frac{1}{2}} \tag{13}$$

where μ and $\Sigma$ represent the mean and variance of the data distribution, respectively. Finally, the training objective function of the entire network is expressed as follows:

$$Loss = Loss_{VAE} + Loss_{DA} \tag{14}$$

(3) Deep Fusion Module

The core objective of the deep fusion module is to fuse latent representations H obtained from multiple views into a unified shared latent representation Z. The fused feature is fed into the clustering module as input to optimize clustering performance.

$$Z = Contact(H(1), H(2), \ldots H(i)) \quad (15)$$

The concatenation operation implemented on latent features H(i) is defined in Formula (15) to facilitate subsequent processing of shared latent representations, where Concat denotes feature concatenation.

Two expansion submodules, namely the up-sampling layer and down-sampling layer, are embedded in the deep fusion module to strengthen the feature representation capability of the shared latent variable Z. To prevent the model from converging to suboptimal solutions with the stacking of multiple fusion layers, residual block structures are integrated into this module. The structural diagram of the module is given in Figure 5.

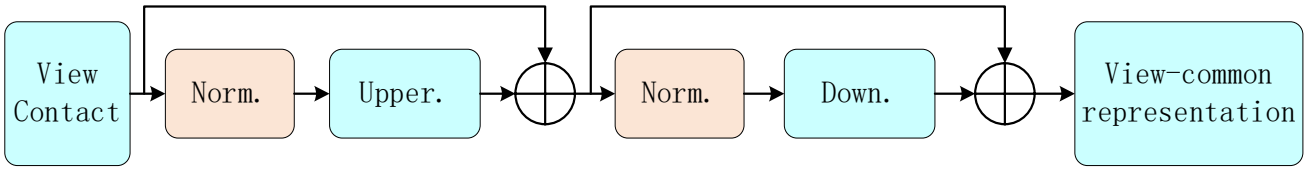



**Figure 5.** Deep fusion module

$$RB(Z) = M(norm(Z)) + Z \quad (16)$$

$$Upper(Z) = M_{reduce}(Dropout(RB(Z))) + Z \quad (17)$$

$$Down(Z) = M_{scale}(Dropout(RB(Z))) + Z \quad (18)$$

Norm (·) denotes a batch normalization operation. Z stands for intermediate embedding of shared view representations. Dropout (·) means neural network nodes are randomly deactivated with probability p. Mscale, and Mreduce represent expansion network and dimension-reduction network, respectively.

(4) Multi-Scale Measurement Fusion Module

The shared view representation Z is obtained after the above procedures. Multi-scale measurement is subsequently performed to calculate data similarity, which is formulated as:

$$Dist_{fused} = \omega_1 * ED + \omega_2 * LB \quad (19)$$

Similarity measurement for power consumers is realized via a weighted combination of Euclidean Distance and LB_Keogh Distance in this paper. Corresponding weights are assigned to the two distance metrics, respectively. Euclidean Distance is adopted to extract the difference features of power loads, while LB_Keogh Distance is utilized to capture load fluctuation characteristics. The weight coefficients are optimized according to clustering results. Final clustering results are obtained by the well-trained network structure together with optimized weights. In Formula (19), ω1 and ω2 denote the weight parameters of Euclidean Distance and LB_Keogh Distance, respectively; ED and LB represent the similarity values calculated from Euclidean Distance and LB_Keogh Distance, and Distfused stands for the final fused distance metric.

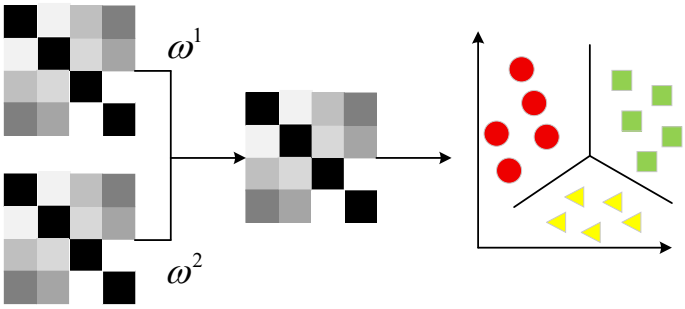



**Figure 6.** Schematic diagram of the clustering process

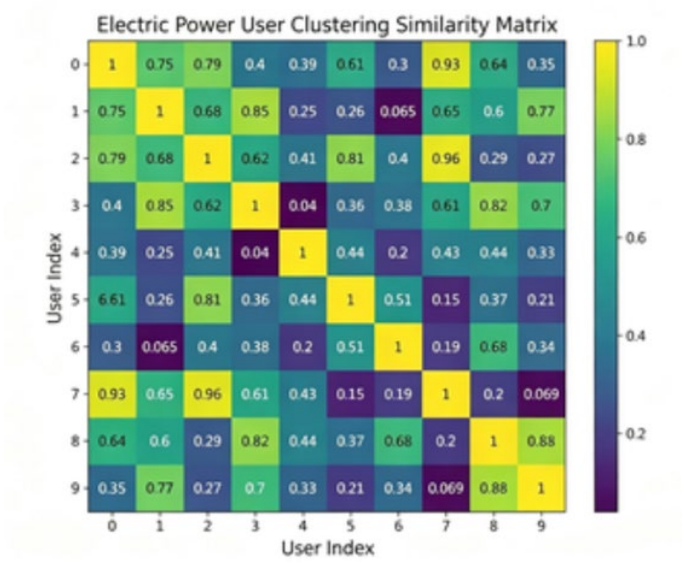



**Figure 7.** User similarity matrix

The clustering procedure is illustrated in Figure 6. A multi-scale distance metric is adopted to calculate the similarity of shared representations, and the fused user similarity matrix is constructed accordingly. Figure 7 presents an example heatmap of a partial clustering similarity matrix for power consumers. The color of each cell corresponds to the pairwise similarity between two users: darker cell color denotes lower similarity, while lighter cell color represents higher similarity. All diagonal elements keep identical values, which indicates perfect self-similarity of each user. In addition, the heatmap is symmetric across the diagonal owing to the symmetric property of the similarity matrix.

### 3.2 Load Characteristic Indexes

The load characteristics of the power system deeply describe various features and attributes of power loads under different production scenarios. Multi-dimensional characteristic indexes are usually required for comprehensive reflection. The industrial power consumption behavior in industrial microgrids is interwoven with complex production rhythms and random fluctuations. In this paper, deep clustering is performed on strong time-series loads through multi-view variational autoencoders. High-dimensional sequences are mapped into low-dimensional latent features. The core representative points of typical power consumption patterns are effectively obtained. Feature extraction of key load fluctuation information is completed. Through in-depth analysis of power consumption behaviors in industrial parks, solid prior feature analysis, and data support are provided for the power consumption behavior identification based on a multi-modal attention mechanism in the next chapter.

Four commonly used clustering evaluation indices are adopted to evaluate the clustering effect: Davies-Bouldin Index (DBI), Dunn Validity Index (DI), Calinski-Harabasz Index (CH), and Silhouette Coefficient (SC). These metrics are extensively utilized in smart meter data profiling to quantify intra-cluster compactness and inter-cluster separation, providing a standard validation framework for clustering effectiveness [22]. These indices comprehensively consider the similarity of each cluster and the separation between different clusters.

$$DBI = \frac{1}{k}\sum_{i=1}^{k} \max_{j\neq i}\left(\frac{S_i+S_j}{M_{i,j}}\right) \tag{20}$$

$$CH = \frac{Tr(B_k)/(k-1)}{Tr(W_k)/(N-k)} \tag{21}$$

$$s(i) = \frac{b(i)-a(i)}{max(a(i),b(i))} \tag{22}$$

$$DI = \min_{1\leq i\leq k}\left\{\min_{j\neq i}\left(\frac{d(C_i,C_j)_{min}}{max_{1\leq i\leq k}\, d(C_l)_{max}}\right)\right\} \tag{23}$$

Where k denotes the total number of clusters; $S_i$ and $S_j$ represent the intra-cluster dispersion of cluster i and cluster j, respectively; $M_{i,j}$ is the distance between the centroid of cluster i and cluster j; Tr(Bk) and Tr($W_k$) stand for the between-cluster variance and within-cluster variance, respectively. a(i) refers to the average distance from sample i to all other samples inside its own cluster, and b(i) denotes the average distance from sample i to all samples in its nearest alien cluster. The value of s(i) ranges from −1 to 1; a value closer to 1 implies denser intra-cluster data distribution and higher inter-cluster separation.

dmin($C_i$, $C_j$) is defined as the minimum sample distance between cluster Ci and cluster $C_j$, while dmax($C_l$) represents the maximum pairwise sample distance within cluster $C_l$.

## 4. Fitting Modeling of Weak Time-Series Loads Based on Adaptive Hybrid Variational Mode Decomposition

### 4.1 Load Modeling Algorithm Based on CEEMDAN-NGO-VMD

Aiming at the characteristics of strong randomness, non-stationarity, and multi-scale feature coupling of weak time-series load data in industrial parks, a “coarse screening-fine classification” hybrid model based on CEEMDAN-NGO-VMD is constructed.

Mode aliasing is effectively suppressed. Refined extraction of complex load components is realized.

Firstly, the CEEMDAN algorithm is adopted to preliminarily decompose the original load sequence. Adaptive white noise is used to reduce the non-stationarity of the sequence and obtain a series of IMF components. Then, the LZC index is introduced to quantify the complexity of each component. High-frequency complex components and low-frequency stable components are adaptively screened by setting a dynamic threshold.

For high-frequency complex components containing aliasing noise, the powerful global optimization ability of the NGO algorithm is used to adaptively optimize the key parameters (mode number K and penalty factor α) of VMD. Secondary fine decomposition is performed.

Finally, the retained low-frequency components and the fine components obtained by secondary decomposition are reconstructed. Complete retention and accurate extraction of load features are realized.

Before performing the NGO algorithm for parameter optimization, initial parameters need to be preset. In this paper, the population size of the Northern Goshawk is set to 30. The maximum number of iterations is 50. The optimization goal is the minimum envelope entropy of VMD decomposition components. The flow chart is shown in Figure 8.

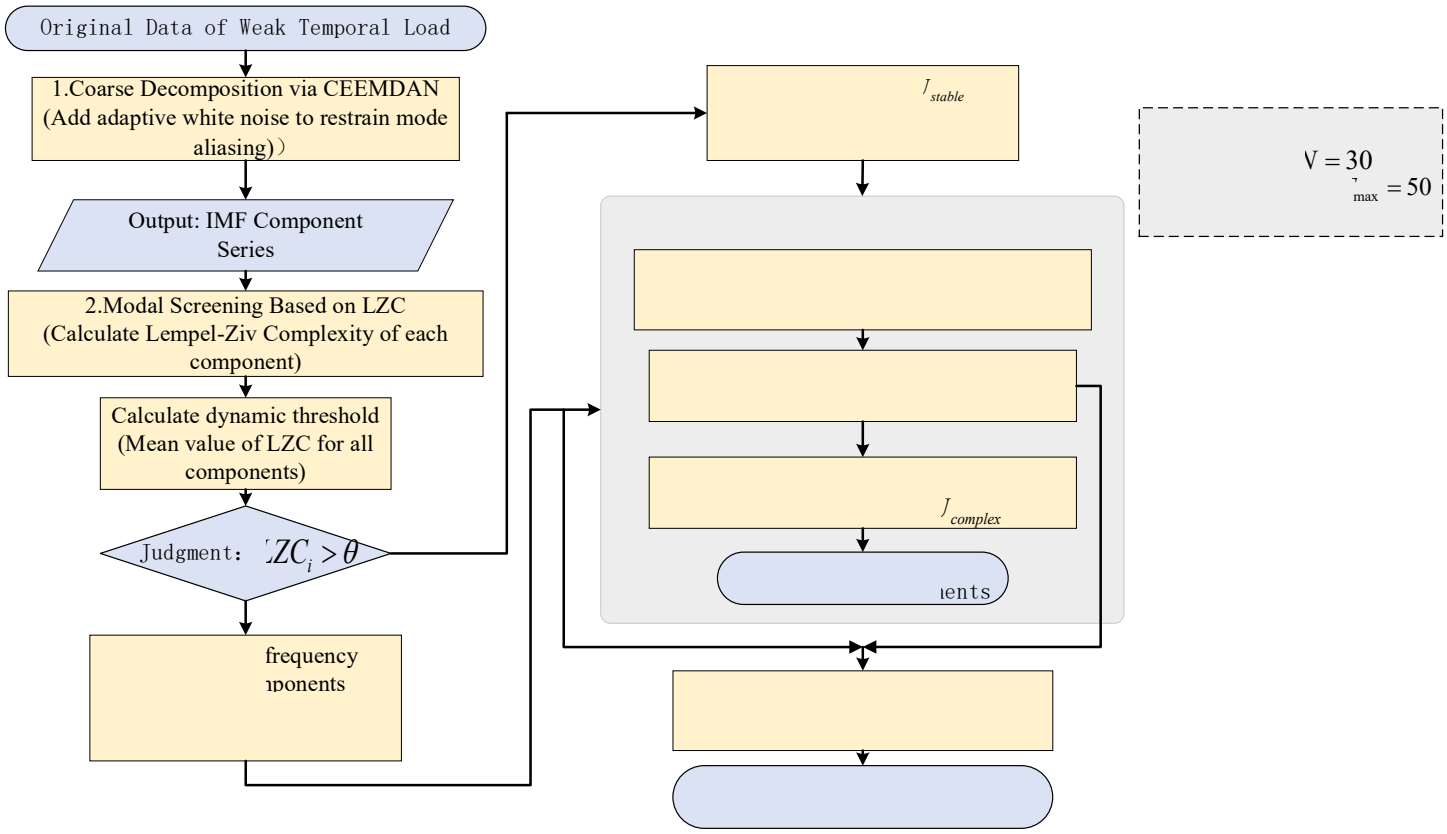



**Figure 8.** CEEMDAN-NGO-VMD flow chart

## 4.2 CEEMDAN Model and Modal Screening

(1) CEEMDAN Model

As an improved method of Empirical Mode Decomposition (EMD) and Ensemble Empirical Mode Decomposition (EEMD), CEEMDAN algorithm is considered as an effective tool for processing nonlinear and non-smooth signals. In the field of power load feature extraction, the CEEMDAN algorithm can adaptively separate IMFs with different physical meanings and time-frequency characteristics from the original load signal. The multi-scale fluctuation modes and internal structure of the load sequence are finely characterized. The discriminative power of feature representation is enhanced.

Firstly, a white noise sequence obeying a standard normal distribution is superimposed on the original signal. Then the EMD method is used to process the noise-superimposed signal to obtain the first group of mode functions:

$$IMF_1^i = EMD\big(x(t) + N_0^i(t)\big), i = 1,2,\dots,N \quad (24)$$

The first intrinsic mode function $IMF_1^i$ is determined by averaging the results obtained by I decompositions.

Subtract the obtained first-order mode function from the original signal to obtain the residual signal r1(t):

$$r_1(t) = x(t) - IMF_1(t) \quad (25)$$

In each decomposition step, white noise is superimposed onto the residual signal rk−1, and EMD decomposition is repeatedly implemented to obtain the k-th group of intrinsic mode functions $IMF_1^i$. The final k-th-order intrinsic mode function IMFk is obtained by averaging the decomposition results derived from multiple noise addition trials.

$$IMF_k = \frac{1}{N}\sum^{i=1}_{N} EMD\big(r_{k-1}(t) + N_0^i(t)\big), k = 1,2,\dots,K \quad (26)$$

Update the residual signal:

$$r_k(t) = r_{k-1}(t) - IMF_2(t) \quad (27)$$

The decomposition terminates when the residual signal rk (t) cannot be further decomposed, or the predefined stopping criterion is satisfied. The complete original signal x(t) can be expressed as the summation of all intrinsic mode functions and the residual component:

$$x(t) = \sum^{k=1}_{K} IMF_k(t) + r_K(t) \quad (28)$$

CEEMDAN is an effective intrinsic mode function extraction technique in signal processing. Adaptive white noise is introduced to suppress mode mixing, which

guarantees the accuracy and integrity of signal decomposition and reconstruction. This algorithm does not require presetting the number of decomposition modes, yet it tends to retain residual high-frequency noise. In this paper, the proposed method is adopted for preliminary coarse decomposition of weakly time-sequential industrial load data. (2) Modal Screening Based on LZC

After preliminary decomposition by CEEMDAN, the original weak time-series load is decomposed into a series of IMFs. Due to the complex composition of weak time-series loads, some IMF components (usually high-frequency components) still have residual noise or mode aliasing, showing strong randomness and non-stationarity. Other IMF components (usually low-frequency components) contain the main trend and periodic characteristics of loads with a relatively regular structure.

In this paper, LZC is introduced as a discriminant index to realize differentiated processing of modal components with different characteristics. Starting from the nonlinear dynamic structure of time series, LZC evaluates the complexity of signals by quantifying the generation rate of "new patterns" in the sequence. This index is more sensitive to signal mutations and non-stationary characteristics of short-time sequences. Structural complexity caused by mode aliasing is effectively identified [23].

The calculation of LZC starts by converting continuous time series into symbolic sequences, followed by counting the occurrence frequency of distinct patterns within the sequence. Let a single IMF component decomposed via CEEMDAN be denoted as H = {h(1),h(2),…,h(N)}. First, compute the mean value $\bar{h}$ of the sequence, then convert H into a binary symbolic sequence S = {s(1),s(2),…,s(N)} based on the following rule:

$$s(i) = \begin{cases} 0, \text{if} h(i) < \bar{h} \\ 1, \text{if} h(i) \geq \bar{h} \end{cases} \quad (29)$$

Scan the symbolic sequence S from left to right and count the number of newly emerging substrings. Define c(N) as the complexity counter initialized to 1. During scanning, the counter increases by 1 whenever a new unseen pattern is detected. This procedure quantifies the generation rate of novel dynamic behaviors in load time series over time.

Normalization is performed on c(N) to eliminate bias induced by varying sequence lengths. For a binary sequence of length N, the upper bound of its complexity is given as:

$$b(N) = \lim_{N\to\infty} c(N) \approx \frac{N}{log_2 N} \quad (30)$$

Hence, the normalized Lempel-Ziv complexity CLZ is defined as:

$$C_{LZ} = \frac{c(N)}{b(N)} = \frac{c(N)\cdot log_2 N}{N} \quad (31)$$

where 0≤CLZ≤1. A larger CLZ value implies more severe mode mixing or noise interference within the corresponding IMF component; conversely, a smaller CLZ indicates stronger deterministic regularity of the time series.

In this paper, an adaptive screening strategy named "coarse pre-screening & precise classification" is proposed. By setting a complexity threshold, all IMF components decomposed by CEEMDAN are categorized into high-frequency complex components and low-frequency stationary components. Assume that K IMF components, denoted as IMF1, IMF2,…, IMFK, are obtained after CEEMDAN decomposition.

The LZC value of each component is calculated according to Formula (3.31), yielding the complexity set:

$$C = \left(C_{lz}(1), C_{lz}(2), \ldots, C_{lz}(K)\right) \quad (32)$$

To accommodate fluctuation characteristics of diverse load datasets, a dynamic threshold θ (equal to the average complexity of all components) is adopted as the classification boundary:

$$\theta = \frac{1}{K}\sum_{i=1}^{K} C_{LZ}(i) \quad (33)$$

If CLZ(i)>θ, the corresponding component features complicated internal composition containing mixed physical modes or intensive noise, which is denoted as $IMF_{complex}$. Such components will be further decomposed elaborately via a parameter self-adaptive optimized VMD algorithm in subsequent steps [24].

If CLZ(i)≤θ, the component mainly reflects a stationary trend or single-period feature of load, marked as $IMF_{simple}$, which is retained directly without participating in follow-up VMD decomposition.

## 4.3 Northern Goshawk Adaptive Variational Mode Decomposition

(1) Northern Goshawk Model

The Northern Goshawk algorithm mainly simulates two stages in the predation process: recognition and attack, pursuit, and escape.

The first stage is the recognition and attack stage. Northern Goshawks search the global space to randomly select prey and launch a rapid attack, aiming to enhance the global exploration ability of the algorithm. The basic equations are as follows:

$$P_i = X_k, i = 1,2,\ldots,N, k = 1,2,3\ldots,i-1,i+1,\ldots,N \quad (34)$$

$$X_{i,j}^{new,P1} = \begin{cases} x_{i,j} + r(P_{i,j} - I_{Xi,j}), F_{Pi} < F_i \\ x_{i,j} + r(X_{i,j} - P_{i,j}), F_{Pi} \geq F_i \end{cases} \quad (35)$$

$$X_i = \begin{cases} X_i^{new,P1}, F_i^{new,P1} < F_i \\ X_i, F_i^{new,P1} \geq F_i \end{cases} \quad (36)$$

where Pi denotes the position of the i-th northern goshawk; FP1 is the objective function value; k is an integer randomly

selected from [1, N] satisfying k≠i; $X_{i,j}^{new,P1}$ represents the updated coordinate on the j-th dimension for the i-th northern goshawk in Phase 1; $X_i^{new,P1}$ is the overall updated position of the i-th northern goshawk; $F_i^{new,P1}$ stands for the objective function value after position update in Phase 1; r and I are random numbers within the range [1, N].

The second phase is the chase-and-escape stage, in which the prey attempts to escape. Benefiting from the high-speed maneuverability of northern goshawks, they can almost capture prey under all practical hunting conditions. By simulating such predatory behavior, the algorithm improves the utilization of local search within the global search space, and its mathematical formulation is expressed as:

$$x_{i,j}^{new,P2} = x_{i,j} + R(2r-1)x_{i,j} \tag{37}$$

$$R = 0.02(1-\frac{t}{T}) \tag{38}$$

$$Xi = \begin{cases} X_i^{new,P2}, F_i^{new,P2} < F_i \\ X_i, F_i^{new,P2} \geq F_i \end{cases} \tag{39}$$

where t denotes the current iteration number, and T stands for the maximum iteration count; $X_i^{new,P2}$ is the updated position of the i-th northern goshawk; $X_{i,j}^{new,P2}$ represents the new coordinate on the j-th dimension of the i-th northern goshawk in Phase 2; $F_i^{new,P2}$ is the updated objective function value of the i-th northern goshawk after the second-phase position update.

(2) VMD

As a time-frequency analysis approach, Variational Mode Decomposition (VMD) decomposes multi-component signals into multiple single-component amplitude-modulated sub-signals. By introducing variational optimization and regularization terms, VMD overcomes inherent drawbacks, including end effect and mode mixing existing in the widely-used Empirical Mode Decomposition (EMD). It possesses rigorous mathematical fundamentals and is more suitable for processing nonlinear and non-stationary signals.

The core principle of VMD lies in the construction and solution of variational problems. As an adaptive and non-recursive decomposition algorithm, it adaptively splits the input sequence into k intrinsic mode functions based on signal frequency-domain characteristics. Its key implementation procedures are listed as follows:

1) Construction of a variational problem. To guarantee each intrinsic mode function is characterized by a specific central frequency and constrained bandwidth, a constraint is imposed: the summation of all decomposed intrinsic mode functions equals the original signal, while the cumulative estimated bandwidth of all modes is minimized. The corresponding mathematical formulation is given below:

$$\min_{\{uk\},\{\omega k\}} \sum_k \| \partial_t[(\delta(t)+\frac{j}{\pi t})u_k(t)]e^{-j\omega kt} \|_2^2 \quad s.t. \sum_k u_k = f \tag{40}$$

where k denotes the iteration count; uk is the k-th decomposed component of VMD; wk represents the central frequency of the k-th modal component; δ(t) is the Dirac delta function; f stands for the weakly time-sequential load series.

2) The constrained variational problem is converted into an unconstrained one by introducing a quadratic penalty factor α and a Lagrange multiplier λ. The resulting augmented Lagrangian function is formulated as:

$$Lu_k, w_k, \{\lambda\} = \alpha\sum_k \| \partial_t[(\delta(t)+\frac{j}{\pi t})u_k(t)]e^{-jekt} \|_2^2 + \| f(t) - \sum_t u_k(t) \|_2^2 + \lambda(t), f(t) - \sum_k u_k(t) \tag{41}$$

3) To solve the unconstrained variational problem, alternating updates and iterations for {u}, {w}, and {λ} are performed via the dual ascent method, and an iterative updating loop is constructed for solving Equation (3.16). The corresponding mathematical expressions are given as follows:

$$w_k{}^{n+1} = \frac{\int_0^\infty |\widehat{u_k}(w)|^2 dw}{\int_0^\infty |u_k(w)|^2 dw} \tag{42}$$

$$\lambda^{n+1}(w) \leftarrow \hat{\lambda}^n(w) + \tau\hat{f}(w) - \sum_{i=1}^{n} u_i^{n+1}(w) \tag{43}$$

$$\sum_k u_k^{n+1} - u_{k2}^{n2}/u_{k2}^{n2} < \varepsilon \tag{44}$$

4) Finally, the solution to the secondary optimization problem is obtained as:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i\neq k}\hat{u}_i(\omega) + \frac{\hat{\lambda}(\omega)}{2}}{1+2\alpha(\omega-\omega_k)^2} \tag{45}$$

The update rule for central frequency is derived as:

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega|\hat{u}_k(\omega)|^2 d\omega}{\int_0^\infty |\hat{u}_k(\omega)|^2 d\omega} \tag{46}$$

Proper parameters [K,α] need to be preset when implementing VMD decomposition. An excessively large K leads to over-decomposition, whereas an undersized K results in insufficient decomposition. Similarly, an overlarge α brings risks of frequency-band information loss, while an excessively small α causes redundant information. Accordingly, the NGO algorithm is adopted to optimize VMD parameters according to signal fluctuation characteristics and complexity.

The optimization flowchart of VMD based on NGO is illustrated in Figure 9. By simulating the predatory behavior of northern goshawks, envelope entropy is taken as the fitness function to calculate the fitness value of each goshawk, determine the optimal objective function value, and save the corresponding optimal position. Iterative updates proceed repeatedly until the termination criterion is satisfied, and the

optimized VMD optimal parameters [K,α] are finally obtained.

For the load components in this paper, the initial population size of the NGO is set to 30, and the maximum iteration number is set to 50 to balance computational efficiency and global search capability.

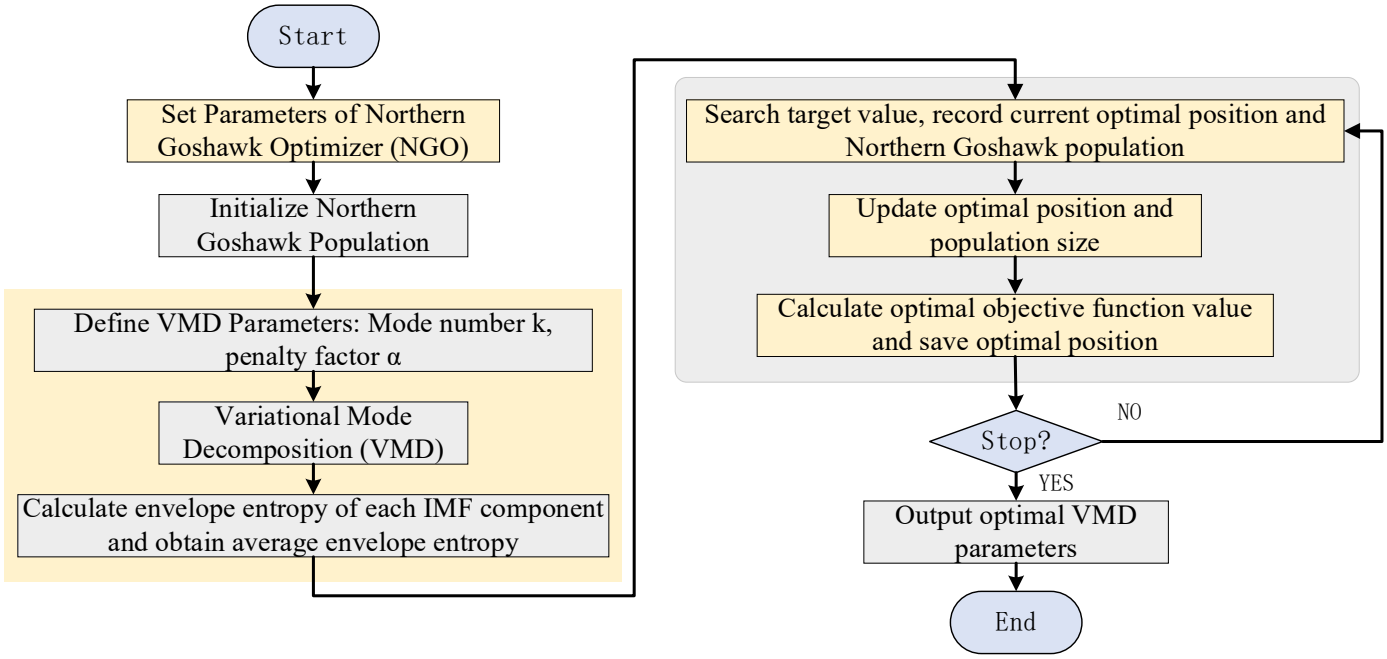



**Figure 9.** NGO-VMD optimization flow chart

## 4.4 Multi-Modal Feature Reconstruction for Microgrid Physical Characteristics

In industrial microgrid scenarios, the original net load sequence P(t) consists of deeply coupled data from rigid industrial production capacity, distributed photovoltaic fluctuation, and frequency regulation scheduling of energy storage systems. According to the spectral distribution characteristics of each IMF component, all IMFs are reconstructed into three physical feature modes in this paper. This approach eliminates dimensional redundancy caused by excessive IMFs after CEEMDAN-NGO-VMD decomposition and provides labeled inputs with definite physical meanings for subsequent electricity consumption pattern identification.

Assume that K valid IMF components are obtained after decomposition. Two frequency split points n1 and n2 (1<n1 <n2<K) are selected based on their normalized central frequencies, to aggregate and classify the K components into high-frequency stochastic mode Ihigh(t), mid-frequency fluctuation mode Imid(t), and low-frequency trend mode Ilow (t). The reconstruction formulas are defined as:

$$I_{high}(t)=\sum_{k=1}^{n_1} IMF_k(t) \qquad I_{high}(t)=\sum_{k=1}^{n_1} IMF_k(t) \tag{47}$$

$$I_{mid}(t)=\sum_{k=n_1+1}^{n_2} IMF_k(t) \tag{48}$$

$$I_{low}(t)=\sum_{k=n_2+1}^{K} IMF_k(t) \tag{49}$$

In this paper, the energy proportion index Er is adopted to quantify the contribution of each reconstructed mode to the overall operating state of the microgrid. For any reconstructed mode Im∈{Ihigh, Imid, Ilow}, its energy percentage relative to the total fluctuation energy of the original net load is calculated as:

$$E_r(I_m)=\frac{\sum_{t=1}^{T} I_m(t)^2}{\sum_{t=1}^{T} P(t)^2}\times 100\% \tag{50}$$

## 5. Case Analysis

To verify the effectiveness of the proposed multi-view VAE deep clustering model and subsequent modal decomposition algorithm, real historical electricity load data of typical industrial consumers (including iron and steel plants, rubber manufacturers, textile mills, cement plants, etc.) from an industrial park in Liaoning Province are selected as case inputs. Considering the periodic characteristics of industrial production, the sampling interval is set to 15 minutes, corresponding to 96 sampling points per user per day; such

sampling resolution achieves a favorable balance between feature retention and computational efficiency.

In data preprocessing and branch segmentation, the average entropy $\bar{E}$ of all users is calculated following the MSE classification criterion defined in Section 3.1. For the clustering experiment in this section, 200 time-series industrial users within the park with three-month measurement data are collected to construct the strong-time-sequential load dataset Dstrong as the input of the multi-view VAE model. Combined with prior planning from the park's marketing department, the preset cluster center number is set as K=4 in the model parameter configuration.

The Euclidean distance weight α and LB_Keogh distance weight β within the fused distance metric matrix are adaptively optimized according to the loss convergence during network training. In the multi-scale fusion module, the initial weights of Euclidean distance and LB_Keogh distance are symmetrically set to α=0.5 and β=0.5 to avoid bias introduced by artificial prior weights. To eliminate bias from artificial weight assignment, a grid search strategy guided by Silhouette Coefficient (SC) maximization is introduced in model training. Dynamic optimization is performed within the range [0,1] with a step size of 0.05 to adaptively determine optimal weight allocation, maximize intra-cluster compactness and inter-cluster separation, and guarantee the clustering results objectively reflect the actual distribution boundaries of four typical industrial load types.

Hardware environment: Windows 10, implemented on MATLAB and Python; Intel i5-8265U @1.60GHz CPU with 12 GB RAM.

To verify the stability of experimental results and avoid accidental conclusions, 5-fold cross-validation is adopted for both the clustering experiment and the decomposition experiment in this study. The dataset is randomly divided into 5 subsets, 4 of which are used for model training and parameter optimization, and 1 for testing. The average value of 5 repeated experiments is taken as the final evaluation result.

(1) Operation Characteristic Analysis of Strong-Time-Sequential Industrial Load Based on MSE Classification

Adopting a multi-view autoencoder framework and multi-scale metric fusion strategy, the multi-view VAE clustering model simultaneously extracts global discrepancy features and local temporal fluctuation features of industrial load, improving the multi-dimensional accuracy of typical mode classification for industrial consumers with strong time-series characteristics. Its outstanding performance in clustering experiments on real industrial park datasets provides high-quality prior labels and reliable data support for the subsequent multi-mode electricity consumption identification model under complex production conditions.

Figure 10 presents continuous 3-day actual load curves of four typical strong-time-sequential industrial consumers, which intuitively demonstrate the temporal characteristics of the input dataset and verify the reliability of the precondition for subsequent multi-view clustering.

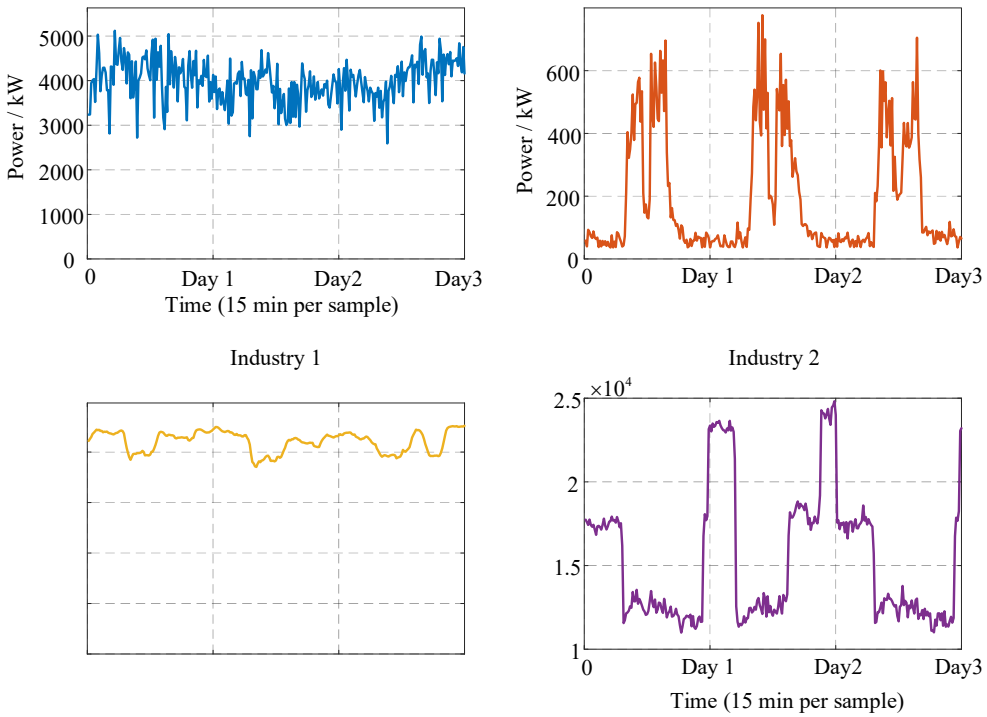



**Figure 10.** Operating characteristics of four typical strong time series industrial loads based on MSE screening

It can be seen from Figure 10 that these load sequences all show high production cycle laws.

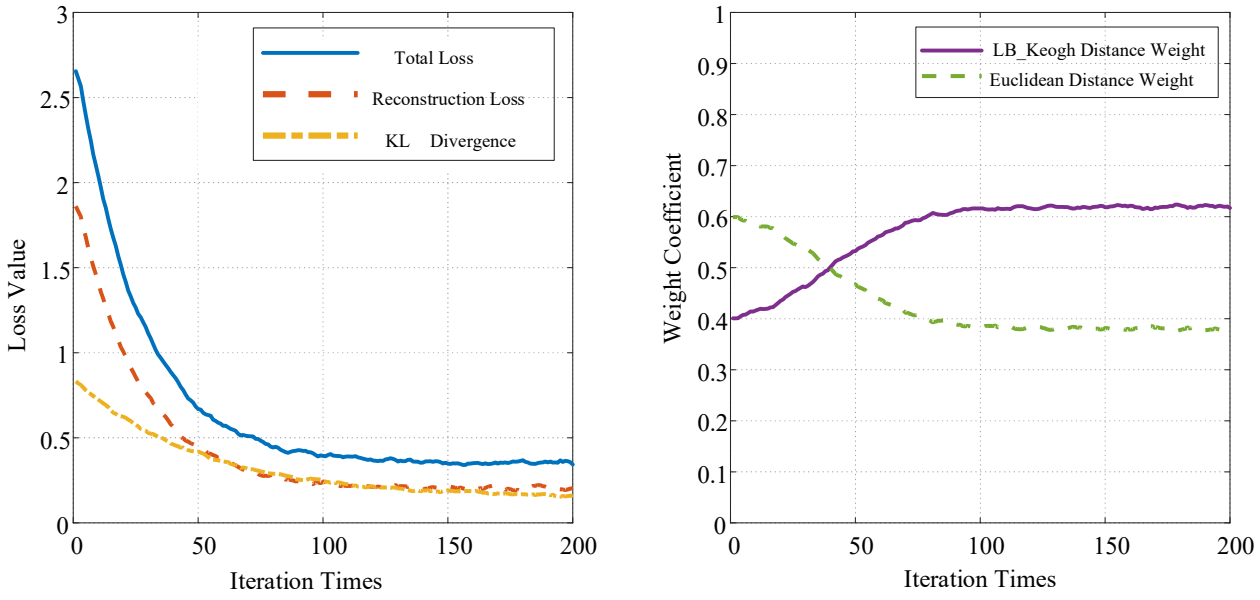
Industry 1 (Iron and Steel Industry): It is characterized by a continuous high-load fluctuation type. Affected by heavy-duty processes such as continuous smelting batches, high-frequency and severe regular fluctuations are shown in specific periods.

Industry 2 (Rubber Manufacturing): It presents a typical daytime discrete alternating law. The nighttime basic load is extremely low. Obvious peak alternations appear during daytime production shifts, which are highly consistent with the daytime scheduling rhythm of the factory.

Industry 3 (Textile Industry): It is characterized by continuous, stable operation. Due to the 24-hour assembly line's high-continuous operation characteristics usually possessed by the textile industry, the load curve shows high continuous stability with minimal daytime fluctuation.

Industry 4 (Cement Industry): It depicts a typical high-energy step-impact type. The power volume is huge. With the start and stop of large equipment such as crushing and calcination, the load curve presents a regular step-like, drastic change.

(2) Multi-View VAE Deep Clustering Model Training and Parameter Optimization Verification.



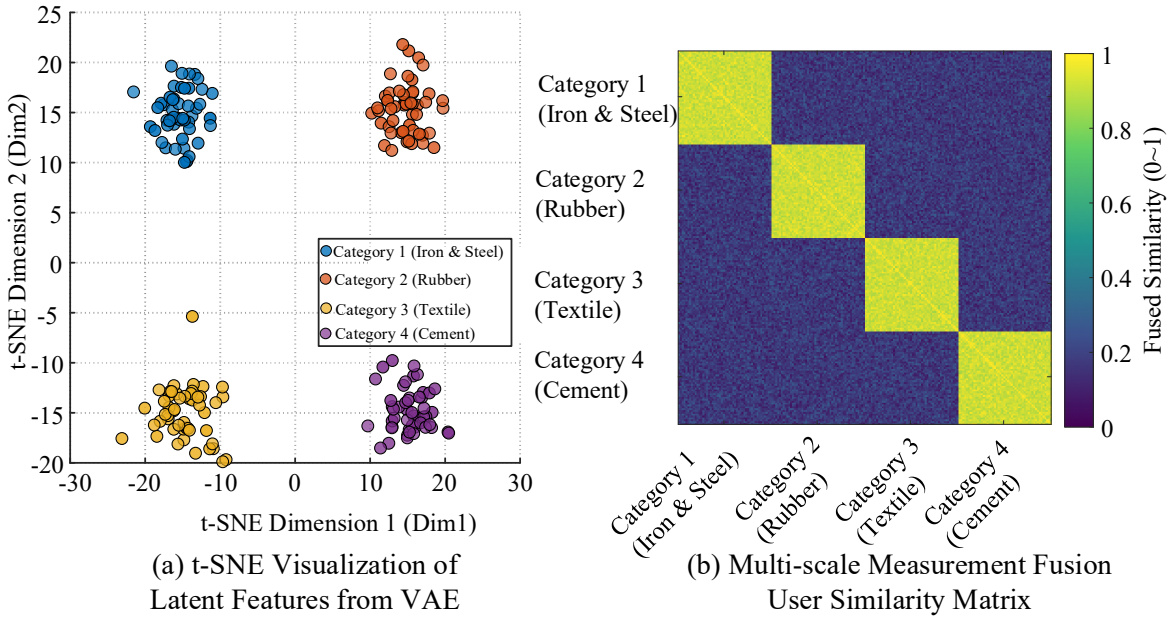

(a) Convergence Curve of Real Training Loss for Multi-view VAE Model

(b) Adaptive Optimization Process of Multi-scale Measurement Fusion Weight

Training and Parameter Optimization Verification of Multi-view VAE Deep Clustering Model

**Figure 11.** Multi-view VAE deep clustering model training and parameter optimization verification

As shown in Figure 11(a), with the increase of iteration times, the total loss, reconstruction loss, and KL divergence of the model show a steady downward trend and tend to converge after 150 iterations. It is indicated that the multi-view autoencoder network successfully learns the high-dimensional nonlinear mapping relationship of industrial loads. No “posterior collapse” phenomenon occurs. Good feature extraction ability and training robustness are shown. Figure 11 (b) shows the dynamic adjustment process of distance weights based on the SC maximization criterion. The weights of Euclidean distance and LB_Keogh distance gradually evolve from the initial 0.5, and finally stabilize at 0.4 and 0.6, respectively.




(a) t-SNE Visualization of Latent Features from VAE

(b) Multi-scale Measurement Fusion User Similarity Matrix

**Figure 12.** Multi-view VAE hidden space feature deconstruction and similarity measurement verification

As shown in Figure 12 (a), after feature extraction and dimensionality reduction by the network, the load data of four types of typical industrial users form aggregated clusters with clear boundaries and no overlap in the two-dimensional space. It can be seen from Figure 12 (b) that the matrix in the user similarity heat map shows an obvious “block diagonal matrix” structure. The highlighted yellow blocks on the main diagonal confirm the extremely high similarity between users of the same type. The dark non-diagonal areas indicate that cross-type power consumption patterns are effectively distinguished.

Three months of data of four types of industrial loads in Liaoning Province are selected for clustering analysis, with one point collected every 15 minutes and 96 points per day. The results are as follows.

(3) Clustering Result Analysis

As shown in Figure 13, Figure 13 (a) is the continuous high-load fluctuation mode: this type of load maintains extremely high power operation all day, but presents severe and high-frequency oscillations. It may represent heavy smelting equipment such as electric arc furnaces in steel plants, which need to keep running continuously. Regular rapid power drops and rebounds are accompanied by the short-cycle process batch alternation of “feeding-melting-tapping”.

Figure 13 (b) is the daytime discrete alternating mode: this load curve shows characteristics synchronized with people's daily routines. The power is extremely low (almost shut down) at night and early morning. Three significant power consumption peaks are formed around key nodes such as 08:00, 13:00, and 19:00. This may represent the production logic and manual intervention characteristics of rubber manufacturing based on shift scheduling.

Figure 13 (c) is the continuous stable operation mode: the clustering center is an almost horizontal straight line with no fluctuation. The baseline fits extremely stably in a specific power range. It may represent the 24-hour continuous production law of “people stop, but machines do not stop” in highly automated assembly lines of textile factories. Except for occasional spikes caused by equipment fine-tuning, the overall equipment is in high-load operation all day.

Figure 13 (d) is the step-impact multi-peak mode: the curve shows a “square wave” step change with large fluctuations, with obvious characteristics of “low during the day and high at night”. It may represent typical high-energy enterprises such as cement plants. To reduce production costs in response to the grid's peak-valley electricity price policy, core heavy equipment such as giant crushers or rotary kilns is often operated at full load during nighttime valley electricity periods. The daily load curve reflects the laws of economic scheduling and step start-stop.

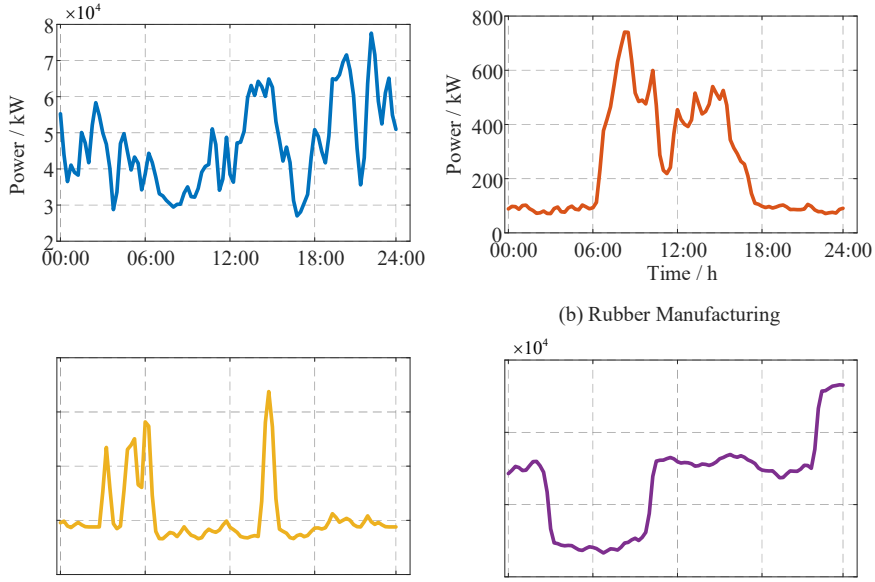



**Figure 13.** Typical industrial daily load power consumption mode extracted by multi-view VAE

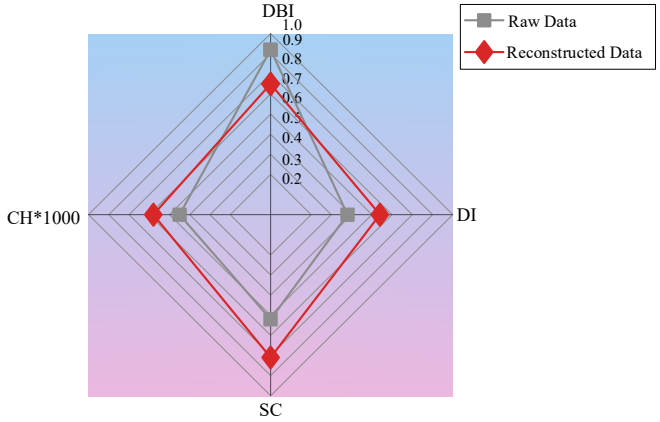



**Figure 14.** Comparison of clustering performance before and after data reconstruction

Table 2. Analysis of clustering results before and after data reconstruction

| | Original Data Clustering | Clustering of Reconstructed Data |
|---|---|---|
| DBI | 0.82 | 0.61 |
| DI | 0.38 | 0.48 |
| SC | 0.52 | 0.87 |
| CH | 4512 | 4857 |

It can be seen from Figure 14 and Table 2 that the DBI of reconstructed data is 25.6% lower than that of original data. The DI, SC, and CH indexes representing inter-cluster separation and intra-cluster compactness are increased by 26.3%, 67.3%, and 7.6%, respectively. Missing and abnormal breakpoints in the original data are found to destroy the time-series continuity of loads, leading to blurred clustering boundaries. The essential fluctuation characteristics of industrial loads are effectively restored by the reconstructed data, making intra-cluster more compact and inter-cluster separation higher. The necessity of missing data reconstruction in improving deep clustering accuracy in the previous chapters is verified.

In this paper, the multi-view VAE deep clustering model is compared with K-means++ algorithm [25], Iterative Self Organizing Data Analysis Techniques Algorithm (ISODATA) [26] and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [27], an established baseline validated in recent smart grid studies for its capacity to decouple cohesive load patterns from irregular noise.

To ensure fair comparison, all baseline algorithms adopt the same optimization criterion (maximizing Silhouette Coefficient, SC) as the proposed method, and the optimal hyperparameters are determined by grid search. The core parameter configurations are as follows:

K-means++: The number of clusters K is searched in the range [2,8] with a step of 1; the k-means++ seeding strategy is adopted for initialization, and the optimal clustering result is selected from 10 repeated runs.

ISODATA: The expected cluster number ranges from 2 to 8, the minimum sample threshold is set to 5, and the maximum number of merge pairs is set to 2.

DBSCAN: The neighborhood radius ε is searched in [0.1,2.0] (normalized data scale) with a step of 0.1, and the minimum sample number MinPts is searched in [3,10] with a step of 1.

All methods share the same input dataset, the same target cluster number K=4 (determined by the park's production planning prior) and the same evaluation index system to ensure equal tuning conditions.

As shown in Figure 15, the proposed method shows the optimal "kite topology".

All evaluation indices reach the optimal level. The three comparison methods are difficult to handle the high-dimensionality and multi-density characteristics of industrial loads. Complex power consumption features are effectively decomposed by multi-view VAE through latent space mapping. Table 3 shows that, compared with DBSCAN, the second-best performing among the three comparison algorithms, the DBI of the proposed method is reduced by 10.3%. DI and SC are increased by 23.1% and 4.8%, respectively. The CH index is also increased by about 6.3%. The superior performance of the proposed algorithm in processing strong time-series industrial loads is verified.

Finally, to verify the effectiveness of the multi-scale measurement fusion module, an ablation experiment of the distance metric mechanism is conducted. As shown in Figure 16, relying only on Euclidean distance is easily disturbed by time-series offset, with a DBI of 0.70. Using only LB_Keogh distance is not sensitive enough to the macroscopic magnitude difference of loads. The dual-distance fusion mechanism adopted in this paper takes into account both global morphological differences and local time-series fluctuations through weight optimization.

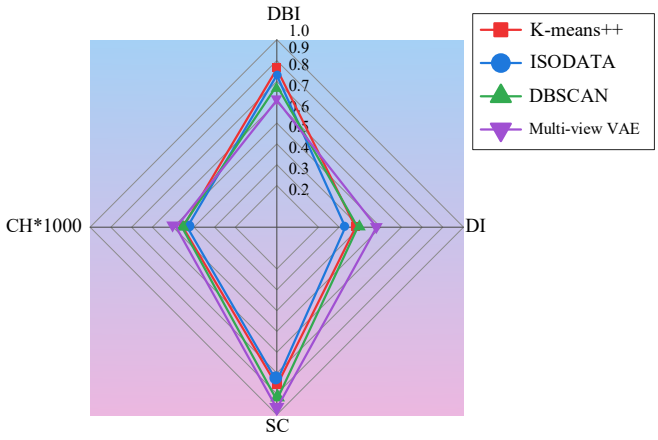



**Figure 15.** Performance comparison of different clustering algorithms

Table 3. Analysis of the results of different clustering algorithms

| | K-means++ | ISODATA | DBSCAN | Multi-view VAE |
|---|---|---|---|---|
| DBI | 0.76 | 0.73 | 0.68 | 0.61 |
| DI | 0.38 | 0.33 | 0.39 | 0.48 |
| SC | 0.76 | 0.74 | 0.83 | 0.87 |
| CH | 4532 | 4323 | 4571 | 4857 |

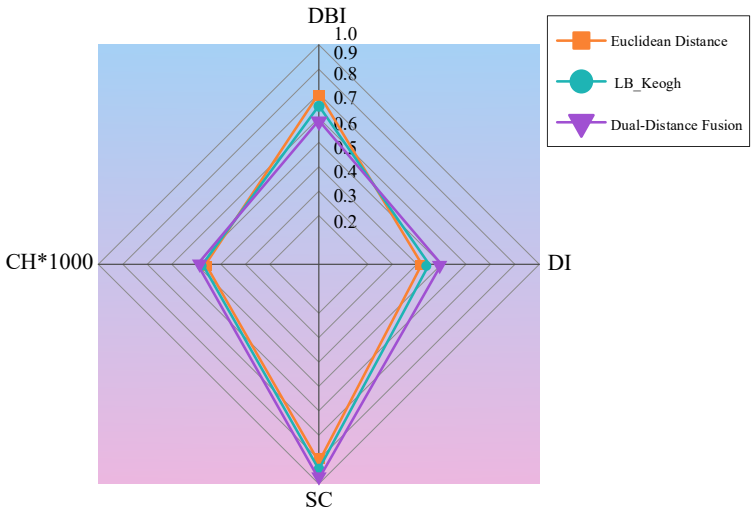



**Figure 16.** Ablation experiment of the similarity measurement mechanism

It can be seen from Table 4 that the dual-distance fusion mechanism reduces DBI by 12.8% and 6.2%, respectively, compared with relying only on Euclidean distance and only using LB_Keogh distance. DI is increased by 14.3% and 6.7%, respectively. SC is increased by 7.4% and 3.6%, respectively. CH is increased by 5.1% and 2.3%, respectively. The dual-distance fusion mechanism adopted in this paper achieves the optimal values in four evaluation indexes. The effectiveness of the fusion module is verified.

Table 4. Analysis of ablation experimental results of the similarity measurement mechanism

| | Euclidean Distance | LB_Keogh | Dual Distance Fusion |
|---|---|---|---|
| DBI | 0.70 | 0.65 | 0.61 |
| DI | 0.42 | 0.45 | 0.48 |
| SC | 0.81 | 0.84 | 0.87 |
| CH | 4621 | 4750 | 4857 |

In the industrial microgrid environment, the access to high-proportion distributed photovoltaics makes the net load curve of industrial users show stronger nonlinearity and multi-peak characteristics. The four types of typical daily power consumption patterns extracted in this paper accurately characterize the energy consumption characteristics of production processes. A high-reliability benchmark label system is provided for subsequent multi-modal power consumption behavior identification in the microgrid context.

To verify the accuracy of the proposed modeling method, loads without obvious periodicity in the park are studied. The measured load data of a certain area in Liaoning is taken as a sample for case verification. The collection interval is 15 minutes, 96 data points per day, for a total of 3 months.

In this paper, the Index of Orthogonality (IO) and RMSE are selected as a comprehensive evaluation system. The performance of different decomposition algorithms in the scenario of complex industrial loads in industrial microgrids is compared.

In an ideal situation, each IMF after signal decomposition should be absolutely orthogonal. The orthogonality index can be used to quantify the degree of independence between decomposed modes, which is defined as follows:

$$IO = \frac{\sum_{j=1}^{K}\sum_{k=1,k\neq j}^{K}\left|\sum_{t=1}^{N} c_j(t)c_k(t)\right|}{\sum_{t=1}^{N} x_{ac}^2(t)} \tag{51}$$

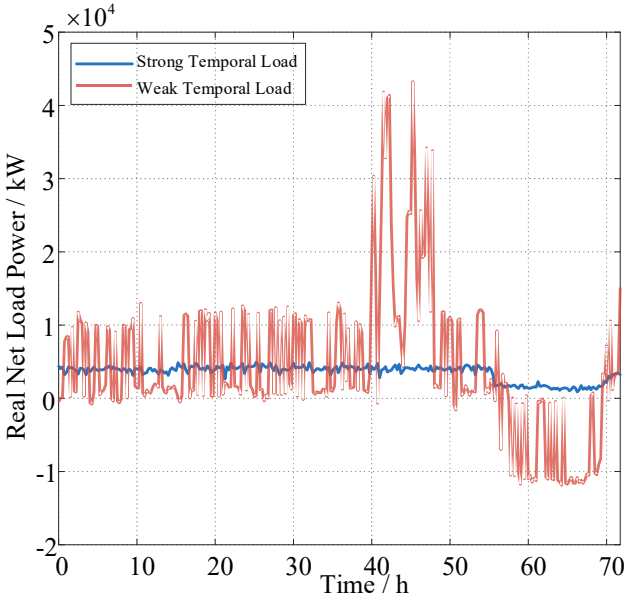
where K is the total number of decomposed modes; N denotes the total sampling points of the load series; cj(t) and ck(t) represent the AC components of the j-th and k-th modes after removing DC baseline, respectively; xac(t) is the AC component of the original load series with mean value eliminated. A value of IO closer to 0 means weaker mutual interference between modal components and stronger capability of the algorithm to suppress mode aliasing.

RMSE quantifies the absolute deviation between the reconstructed signal superimposed by all decomposed modes and the original signal, serving as a core indicator to check severe energy leakage or information loss of the algorithm. Here, the true value refers to the original load sequence, while the predicted value denotes the reconstructed series obtained by summing all decomposed modes and global residual error. A smaller RMSE indicates better energy conservation performance of the algorithm.

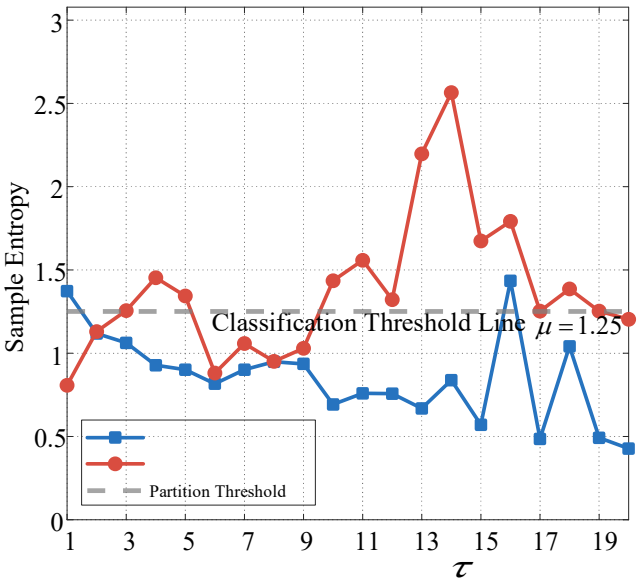
(1) Operation Characteristic Analysis of Weak-Time-Sequential Industrial Load Based on MSE Classification

Within industrial microgrid, user-side net load is no longer a single industrial power consumption curve. Due to grid integration of high-penetration randomly distributed clean energy output and frequent charge-discharge regulation of energy storage equipment, deep coupling among source, grid, and storage heavily masks the inherent power consumption rules of partial industrial loads. Therefore, temporal complexity classification is performed on the load before feature extraction.

Multiscale Entropy (MSE) is introduced to measure the complexity of industrial user load series. Sample entropy effectively evaluates nonlinear dynamic characteristics and self-similarity of time series, while multi-scale coarse-graining reveals internal evolution rules of load across different time scales.




(a) Comparison of Real Load Waveforms for Strong and Weak Time Series




(b) Complexity Measurement of Real Load Based on MSE

**Figure 17.** Comparison of strong and weak time series industrial load characteristics and MSE curves

As shown in Figure 17(a), strong time-series loads have small fluctuations and clear boundaries of production start-stop. Weak time-series loads are affected by both clean energy random fluctuations and load-side sudden changes. The waveform shows burrs and disorder. It can be seen from Figure 17(b) that the MSE curve of strong time-series loads basically remains below the classification threshold line with the increase of the scale factor, indicating high predictability and self-similarity. The MSE values of weak time-series loads are relatively high at all scales. A large number of irregular random modes and complex dynamic characteristics are contained within the sequence.

(2) CEEMDAN Preliminary Decomposition Results

As shown in Figure 18, the top is the original weak time-series industrial load sequence with a messy waveform and many sudden impacts. After preliminary decomposition by CEEMDAN, the original signal is stripped into multiple IMF components with frequencies from high to low. It can be seen from the waveform that low-frequency components gradually show a smooth periodic trend. Dense non-stationary oscillations are still found in the first few high-frequency IMF components. It is indicated that a single CEEMDAN algorithm is difficult to fully decompose complex industrial loads. The high-frequency part still has a certain degree of mode aliasing and noise pollution. Quantitative indicators are required for further screening and refined processing.

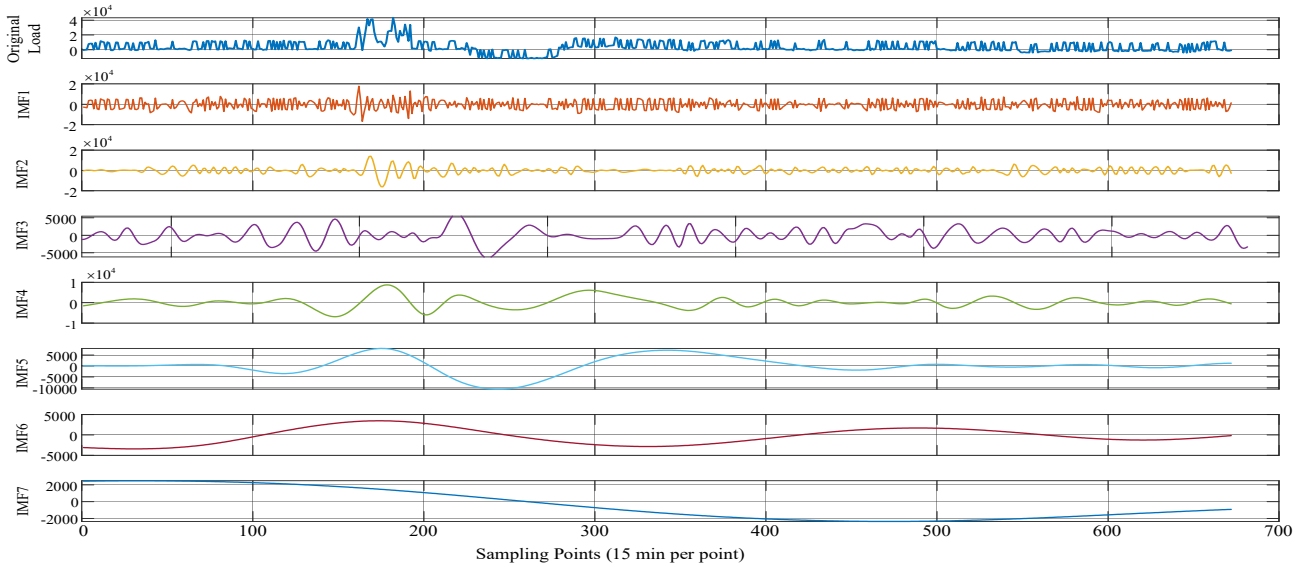



**Figure 18.** The original sequence of weak time series industrial load and the preliminary decomposition results of CEEMDAN

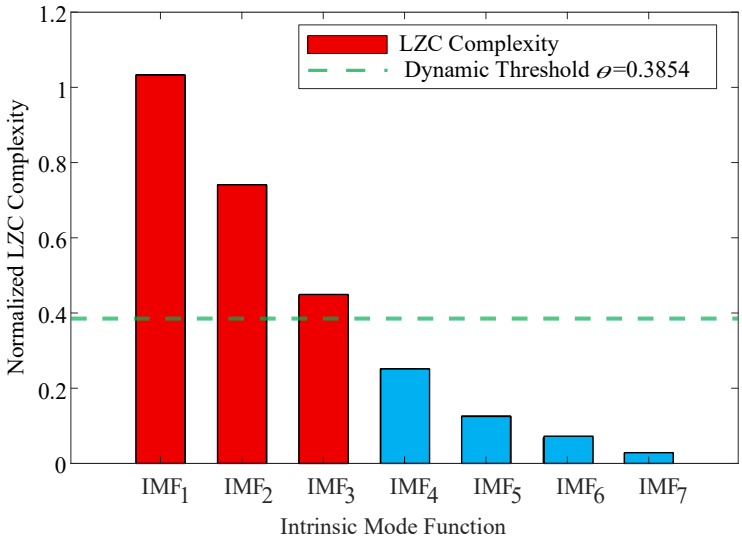



**Figure 19.** Lempel-Ziv complexity and adaptive threshold screening of each IMF component

As shown in Figure 19, the dynamic complexity threshold after decomposition is θ=0.3854. The complexity of IMF1, IMF2 , and IMF3 is significantly higher than the threshold, which are high-frequency complex components. It indicates that these high-frequency modes contain obvious mode aliasing and severe random noise interference. Direct reconstruction or feature extraction using these components will degrade the accuracy of subsequent models. In this paper, the three components are classified into the set requiring refined decomposition and fed into the NGO-VMD model for secondary fine decomposition. The LZC values of IMF4 to IMF7 are all below the threshold line, corresponding to low-frequency stationary components. With regular waveform structures, such components mainly reflect the macroscopic periodic evolution and stable trend characteristics of load series; hence, they can be directly reserved and involved in final signal reconstruction without secondary decomposition.

(3) Analysis on Secondary Fine Decomposition Results of NGO-VMD

Figure 20 compares the fitness convergence curves of NGO, classical Particle Swarm Optimization (PSO), and Grey Wolf Optimizer (GWO) [28] when optimizing the core VMD parameters [K,α]. Under identical initial conditions, all three intelligent algorithms converge to the same minimum fitness value with iterative optimization. This proves that a unique globally optimal parameter solution exists within the preset boundary for the current high-frequency load components.

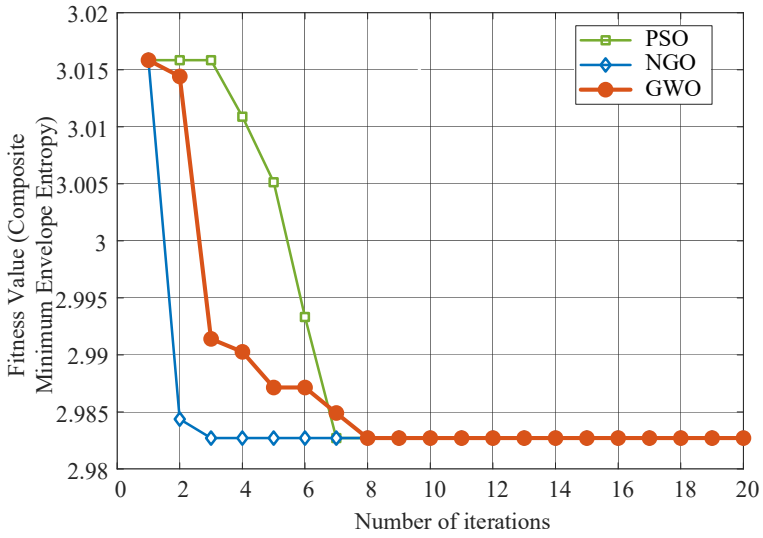



**Figure 20.** Comparison of fitness convergence curves of VMD optimized by multi-intelligent algorithms

As illustrated in Figure 20, obvious discrepancies exist among the three algorithms in terms of convergence speed and optimization searching process.

The fitness curves of PSO and GWO exhibit varying degrees of flat stagnation in the early iteration stage, which reveals that these two algorithms are prone to falling into local optimum when coping with complicated search spaces. By contrast, the proposed NGO algorithm possesses superior global search capability and better ability to escape from local optimum. Its fitness curve drops rapidly at the initial iterations and reaches stable convergence with the minimum iteration counts. The comparative results verify that the NGO algorithm can quickly acquire the optimal VMD parameter combination matching the current high-frequency components and effectively shorten the optimization computation time in the feature extraction phase of the model.

The optimal parameters optimized by the NGO are assigned to the VMD model to conduct secondary decomposition on the high-frequency complex components screened via LZC, and the corresponding results are displayed in Figure 21. Compared with the preliminary decomposition waveforms containing glitches in Figure 18, the seven VMF components extracted after secondary refined decomposition by NGO-VMD feature smoother and more regular waveforms, and the local random abrupt impulses are effectively suppressed.

(4) Time-Frequency Domain Decomposition Result Analysis of Weak Time-Series Loads by CEEMDAN-NGO-VMD.

As shown in Figure 22, the model proposed in this paper reconstructs 8 effective modes. The left time-domain waveform is the evolution process of the industrial load from slight oscillation to regularity. The right spectrogram is the frequency-domain quality of decomposition. It can be seen that the spectral peaks of each mode are distributed in their independent frequency bands. The crossover and overlap between them are reduced to an extremely low level. This feature of clear frequency band boundaries verifies that the proposed algorithm can effectively suppress the common mode aliasing and frequency band mutual interference in traditional decomposition methods.

To ensure equal tuning of all decomposition methods, the parameter configurations and optimization objectives of baseline algorithms are consistent with the proposed framework as follows:

CEEMDAN: The adaptive white noise amplitude coefficient is set to 0.2, the ensemble number is set to 100, and the stopping criterion is that the number of extreme points of the residual signal is less than 2, which follows the standard configuration of classical literature.

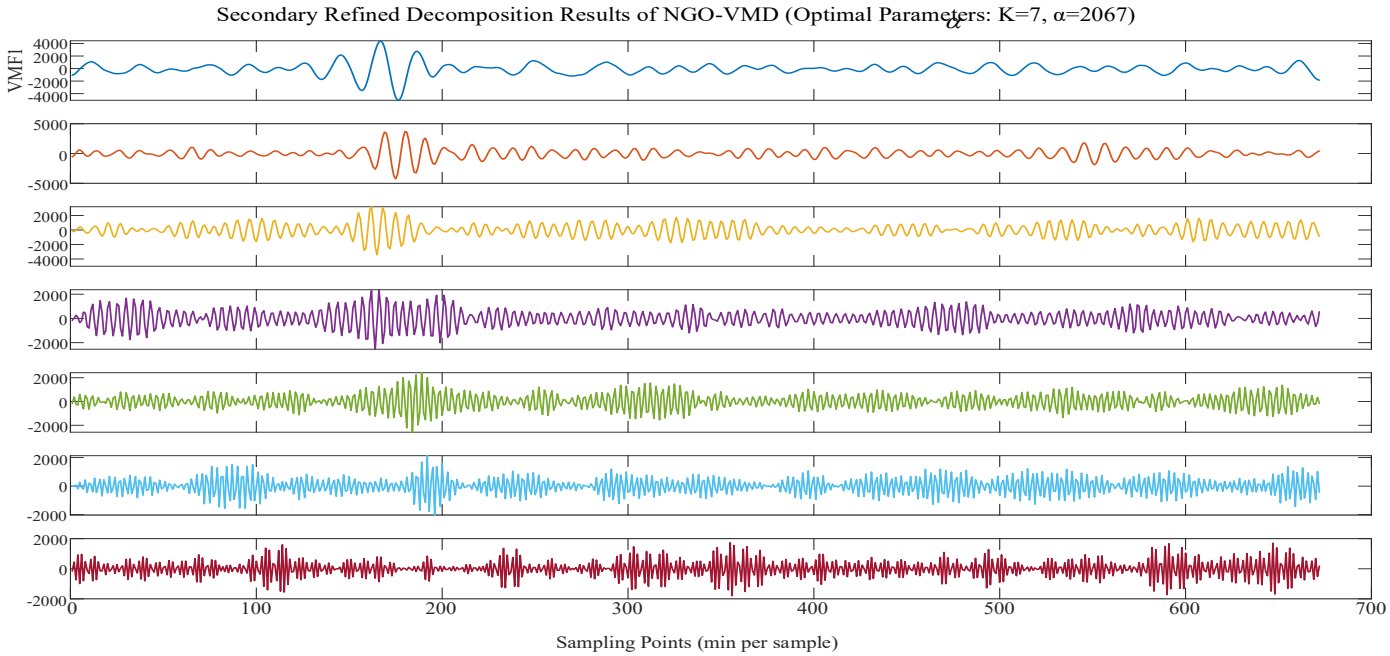



**Figure 21.** NGO-VMD secondary fine decomposition result diagram

NGO-VMD: The NGO optimizer adopts exactly the same parameters as the proposed method (population size = 30, maximum iterations = 50); the VMD mode number K is searched in [4,10], and the penalty factor α is searched in [1000,3000], with the minimum envelope entropy as the optimization objective.

It can be seen from Table 5 that although the CEEMDAN-based algorithm [29] can achieve extremely low energy loss reconstruction, its IO is relatively high, reaching 0.2748, indicating obvious mode aliasing defects. The NGO-VMD algorithm [30] has better orthogonality than the CEEMDAN algorithm, but its RMSE is higher. In contrast, the IO of the proposed CEEMDAN-NGO-VMD shunting architecture is only 0.1512, 45% lower than that of the CEEMDAN-only method, and about 29.7% lower than that of the NGO-VMD-only algorithm. Although the introduction of intelligent optimization increases about 25 seconds of offline calculation time, it greatly improves the purity of time-frequency features. A more reliable benchmark label is provided for subsequent deep neural network load power consumption behavior identification.

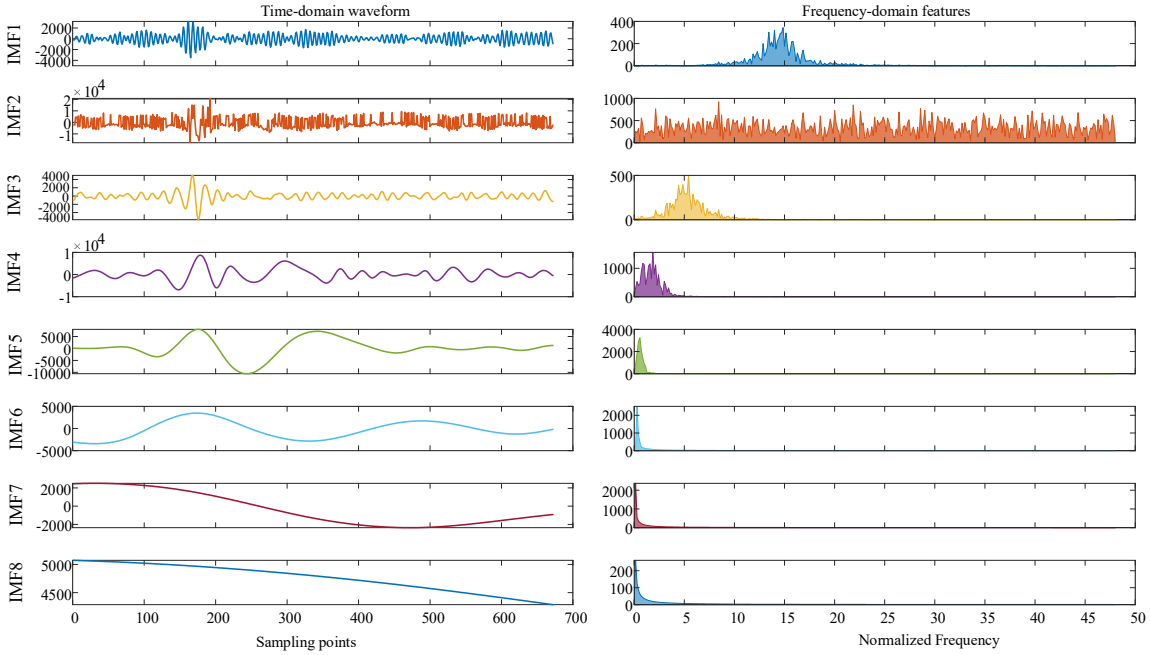



**Figure 22.** Time-frequency domain decomposition results of weak time series load based on CEEMDAN-NGO-VMD

Table 5. Performance evaluation analysis of different signal decomposition algorithms

| Methods | IO | RMSE | Computation Time (s) |
|---|---|---|---|
| CEEMDAN | 0.2748 | 1.12×10-12 | 0.59 |
| NGO-VMD | 0.2150 | 1.66×10-12 | 5.73 |
| CEEMDAN-NGO-VMD | 0.1512 | 3.25×10-13 | 25.68 |

As illustrated in Figure 23, with two frequency dividing boundaries n1 and n2, the heavily mixed and non-stationary original net load is decomposed into three physical modes with distinct heterogeneous characteristics:

Low-frequency Trend Mode $I_{low}$ (mapping rigid production scheduling baseline): This mode features extremely smooth waveforms after eliminating interferences from the external environment and dispatching, accounting for 25.6% of total energy. From the physical mechanism, it precisely characterizes the fundamental energy consumption evolution of industrial users constrained by production procedures and shift schedules.

Mid-frequency Fluctuation Mode $I_{mid}$ (mapping source-grid-load-storage interactive attributes): As the dominant mode reflecting industrial microgrid features, it contains 10.4% fluctuation energy. Its waveform presents violent intraday oscillation and continuous ups and downs in specific periods, which originates from output volatility of distributed photovoltaic and wind power, as well as daily charge-discharge cycles of energy storage systems for peak-valley price arbitrage and photovoltaic fluctuation suppression.

High-frequency Random Mode $I_{high}$ (mapping transient impulse and measurement noise): Dense glitches and high-frequency oscillations exist in its waveform with an energy proportion of 1.5%. Despite the low energy share, it carries vital transient physical information, including white noise from high-precision measurement systems and nonlinear physical impulses induced by the instantaneous startup-shutdown of high-power heavy equipment such as rolling mills and arc furnaces inside the industrial park.

In summary, the single one-dimensional net load series is extended into a three-dimensional feature tensor $Ij \in \{I_{high}, I_{mid}, I_{low}\}$ with definite physical meanings via frequency-band reconstruction. Defined as independent microscopic

fluctuation modes, the tensor is fed together with macroscopic periodic modes extracted from multi-view VAE into the multimodal attention network in Chapter 4 to promote the accuracy of the subsequent deep learning identification model.

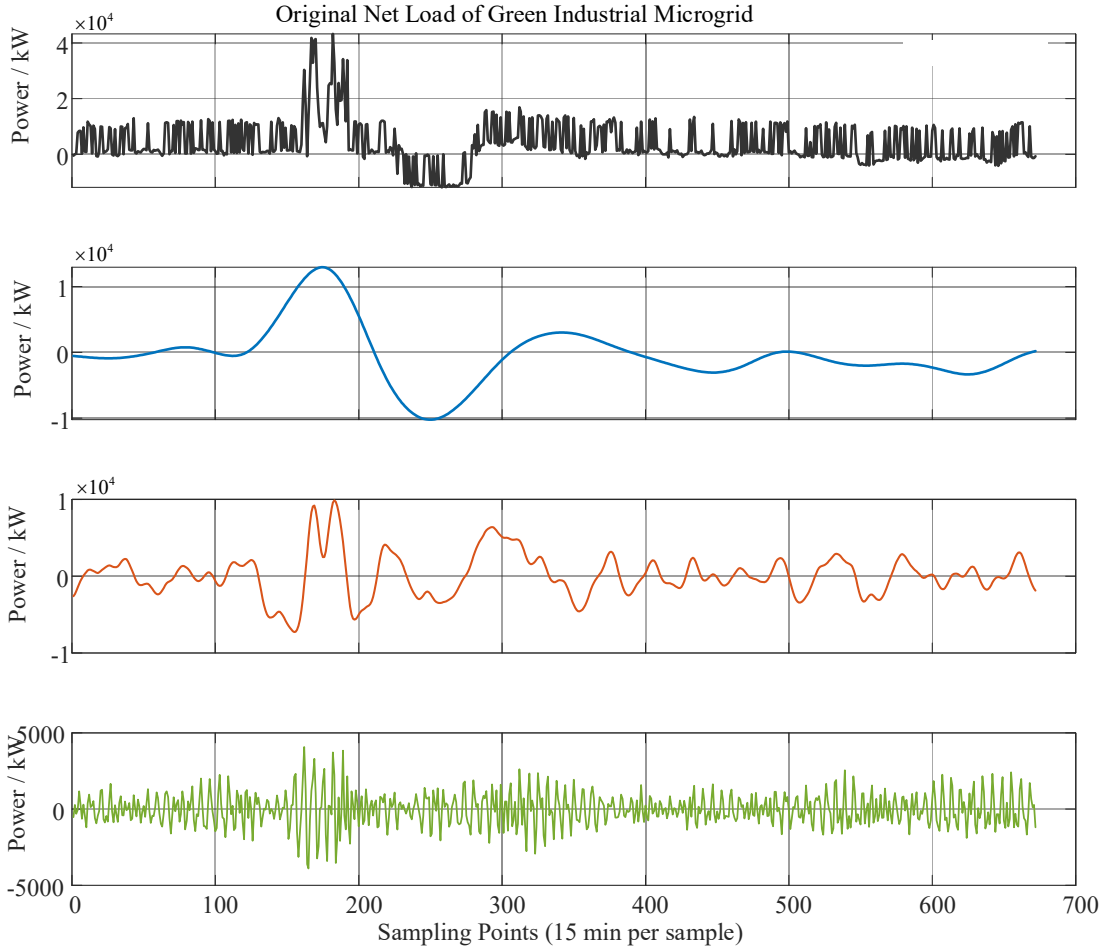



**Figure 23.** Decoupling results of multi-modal characteristics of microgrid net load

## 6. Conclusion

This paper focuses on the shunt synergy analysis and feature extraction of industrial loads under industrial microgrid scenarios. Firstly, multi-scale entropy (MSE) is introduced to quantify the dynamic complexity of load series, realizing refined classification of strongly and weakly time-sequential loads. Secondly, for strongly time-sequential loads, a multi-view VAE deep clustering model embedded with adaptive dual distance metrics is constructed. Finally, for weakly time-sequential loads, a hybrid decomposition framework of "coarse screening & refined decomposition" based on CEEMDAN-NGO-VMD is proposed.

Experimental results demonstrate that the proposed method outperforms existing algorithms in feature decoupling and aliasing suppression performance. In terms of clustering metrics, the Davies-Bouldin index (DBI) declines to 0.61, and the Silhouette Coefficient (SC) rises to 0.87. Compared with DBSCAN (the suboptimal algorithm among the three benchmark approaches), the proposed method reduces DBI by 10.3%, improves DI and SC by 23.1% and 4.8% respectively, and boosts the Calinski-Harabasz index (CH) by approximately 6.3%. Near-lossless signal reconstruction is achieved in mode decomposition with the RMSE as low as 3.25×10−13 and the index of orthogonality (IO) reduced to 0.1512, which is 45% and 29.7% lower than that of standalone CEEMDAN and standalone NGO-VMD, respectively. Experimental results based on measured data from an industrial park in Liaoning show that the categorized processing strategy can effectively eliminate feature aliasing induced by single modeling in the feature extraction of source-load tightly-coupled industrial microgrid loads, and achieves superior performance on this high-noise and non-stationary industrial load dataset. Further verification in multi-region and multi-type industrial microgrid scenarios can be carried out in future work. The extracted high-purity time-

frequency features and high-reliability benchmark typical patterns optimize the input quality of the subsequent dual-attention multimodal fusion network and improve the accuracy and generalization of the electricity consumption behavior identification model.

## Author Contribution

Conceptualization, S.Y. and Y.Z.; methodology, S.Y.; software, S.Y. and L.L.; validation, S.Y., L.L., Z.T., Q.W. and Y.Z.; formal analysis, S.Y.; investigation, L.L., Z.T. and Q.W.; resources, Y.Z.; data curation, S.Y. and Z.T.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y. and Y.Z.; visualization, S.Y. and Q.W.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

## Data Availability Statement

The historical monitoring load data utilized to support the findings of this study were supplied by the industrial microgrid administration of Liaoning Province under license. Due to commercial privacy restrictions and operational security policies of the participating industrial users, the raw dataset cannot be made publicly accessible. However, the pre-processed time-series feature samples, anonymized simulation data, all algorithm implementation codes and simulation parameter configurations are available from the corresponding author upon reasonable academic request, to support the reproducibility of the research.

## References


[1] Khodayar M, Liu G, Wang J, et al. Deep learning in power systems research: A review. CSEE Journal of Power and Energy Systems. 2021;7 (2):209-220.

[2] Zhou M, Kong W, Hu F, et al. OCAE-based feature extraction and cluster analysis of high-energy-consuming plant loads. Computers and Electrical Engineering. 2024;119:109579.

[3] Eren Y, Küçükdemiral İ. A comprehensive review on deep learning approaches for short-term load forecasting. Renewable and Sustainable Energy Reviews. 2024;189:114031.

[4] Reddy ASS, Pachori RB. Adaptive Multi-Resolution Dynamic Mode Decomposition for Non-Stationary Signal Analysis. IEEE Signal Processing Letters. 2026;33:998-1002.

[5] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy. 2012;42 (1):68-80.

[6] Li K, Ma Z, Robinson D, et al. Identification of typical building daily electricity usage profiles using Gaussian mixture model-based clustering and hierarchical clustering. Applied Energy. 2018;231:331-342.

[7] Yang D, Lue S, Wei J, et al. Detection and Classification of Power Quality Disturbances Based on Improved Adaptive S-Transform and Random Forest. Energies. 2025;18 (15):4088.

[8] Motlagh O, Berry A, O'Neil L. Clustering of residential electricity customers using load time series. Applied Energy. 2019;237:11-24.

[9] Cui J, Qu X, Lv C, et al. Optimized modal decomposition techniques for robust leakage detection in noisy environments: A comparative study. Measurement. 2025;252:117390.

[10] Eskandari H, Imani M, Moghaddam M P. A deep residual network integrating entropy-based wavelet packet ensemble model for short-term electrical load forecasting. Energy. 2025;314:134168.

[11] Zhang Y, Zhang J, Yao G, et al. Method for Clustering Daily Load Curve Based on SVD-KICIC. Energies. 2020;13 (17):4476.

[12] Han H, Peng J, Ma J, et al. Research on load forecasting based on CEEMDAN SE VMD and SelfAttention TCN fusion model. Scientific Reports. 2025;15 (1):14530.

[13] Li S, Zhu X, Zhou D. Power quality disturbance signal denoising and detection based on improved DBO-VMD combined with wavelet thresholding. Electric Power Systems Research. 2025;238:111193.

[14] Zamora J, Sublime J. An Ensemble and Multi-View Clustering Method Based on Kolmogorov Complexity. Entropy. 2023;25 (2):371.

[15] Lu F, Ciu X, Xing J, et al. Electricity Load Profile Characterisation for Industrial Users Based On Normal Cloud Model and iCFSFDP Algorithm. IEEE Transactions on Power Systems. 2023;38 (4):3799-3813.

[16] Cai M, Zheng Y, Peng Z, et al. Research on load clustering algorithm based on variational autoencoder and hierarchical clustering. PLOS ONE. 2024;19 (6):e0303977.

[17] Zhang H, Zhou M, Chen Y, et al. Short-term power load forecasting for industrial buildings based on decomposition reconstruction and TCN-Informer-BiGRU. Energy and Buildings. 2025;347:116317.

[18] Yang Y, Li C, Qin Y, et al. Clustering analysis of daily load curve based on convolutional variational autoencoder. Modern Electric Power. 2023;40 (5):770-778.

[19] Dehghani M, Hubálovský Š, Trojovský P. Northern Goshawk Optimization: A new swarm-based algorithm for solving optimization problems. IEEE Access. 2021;9:162059-162080.

[20] Greer C, Liu X A, Best R. Clustering Smart Meter Data for Residential Load Profile Analysis. In: 2025 60th International Universities Power Engineering Conference (UPEC); 2025; London, United Kingdom. London: IEEE; 2025. p.1-5.

[21] Huang F, Zhang X, Li C, et al. Multimodal Network Embedding via Attention based Multi-view Variational Autoencoder. In: 8th ACM International Conference on Multimedia Retrieval (ACM ICMR); 2018 Jun 11-14; Yokohama, Japan. 2018.

[22] Chen Y, Hou H, Zhang Z, et al. Review on Smart Meter Data Clustering and Demand Response Analytics. In: 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC); 2020; Nanjing, China. Nanjing: IEEE; 2020. p.1-6.

[23] Kaspar F, Schuster H G. Easily calculable measure for the complexity of spatiotemporal patterns. Physical Review A. 1987;36 (2):842-848.

[24] Dragomiretskiy K, Zosso D. Variational mode decomposition. IEEE Transactions on Signal Processing. 2014;62 (3):531-544.

[25] Yang M, Hussain I. Unsupervised Multi-View K-Means Clustering Algorithm. IEEE Access. 2023;11:13574-13593.

[26] Jain A K. Data clustering: 50 years beyond K-means. Pattern Recognition Letters. 2010;31 (8):651-666.

[27] Mishra A K, Das B. A Novel Density Based Clustering Approach for Electricity Theft Detection. IEEE Transactions on Industry Applications. 2025;61 (4):5537-5548.

[28] Mirjalili S, Mirjalili S M, Lewis A. Grey Wolf Optimizer. Advances in Engineering Software. 2014;69:46-61.
[29] Tian L, Lao C, Yang Z. Load forecasting based on CEEMDAN-SE-xLSTM-Transformer. The Journal of Supercomputing. 2026;82 (1):40.
[30] Xu F, Jing R, Zhang Y, et al. High accuracy key feature extraction approach for the non-stationary signals measurement based on NGO-VMD noise reduction and CNN-LSTM. Measurement Science and Technology. 2024;35 (1):015031.