International Journal of Power and Energy Systems, Vol. 46, No. 10, 2026
An Open Access Paper

SMART GRID EMERGENCY POWER PROTECTION DATA ANALYSIS METHOD BASED ON AI ENGINE TECHNOLOGY

Di Huang,* Liangrong Xu,* Bin Zhou,* Chaoyue Zhu,* Weiyan Zheng,* and Xingping Yan*

Abstract

This study proposes an enhanced analysis framework based on an artificial intelligence engine, aiming to address the challenges of highly complex data and dynamic changes in user behavior in emergency power protection scenarios of smart grids. This method incorporates a multi-stage data processing flow that includes anomaly detection using isolated forests, data repair via cubic spline interpolation, and behavioral decoupling analysis, which combines information entropy features with spectral clustering. The experimental results showed that the proposed method in the study achieved a classification accuracy rate of 91.4% through kernel space transformation and density peak adaptation. The case analysis revealed significant behavioral insights. Cluster I exhibited characteristic dual peaks (0:00-8:00 and 18:00-24:00 hours) as well as daytime gas slots (8:00-18:00 hours), which precisely matched the three-shift manufacturing operation. On the contrary, Cluster V showed the double peaks of the office (9:00-11:00 and 15:00-17:00 hours), reflecting the working rhythm of the administrative agency. The research found that among the five power consumption patterns, the classification accuracy rate reached 91.4%, and the overall processing efficiency was 37.6% higher than that of the traditional method. This research provides feasible data analysis support for dynamic demand response and stable operation in smart grids.

Key Words

AI engine; Isolated forest; Spline interpolation; Data fusion; Information entropy; Spectral clustering

* Zhejiang Dayou Industrial Co., Ltd. Hangzhou Science and Techology Development Branch, China; euyjqj@163.com, vwclqq@163.com, hzdykj123456@126.com, xuchangle8887@163.com, zpsigb@163.com, liangrongx@mailpo.uu.me,
Corresponding author: Chaoyue Zhu

(DOI: 10.2316/J.2026.203-0669)

1. Introduction

The power industry is undergoing a profound market-oriented transformation due to the accelerated advancement of smart grid technology [1]. As primary stakeholders in energy consumption, end-users are playing an increasingly active role in tariff formulation mechanisms to protect their interests [2]-[3]. Tariff structures are pivotal economic indicators in the power sector. Their determination is influenced by factors related to generation, transmission, and consumption processes [4]. Electricity pricing is dynamically determined by systematically integrating spatio-temporal parameters from operational scenarios and their corresponding energy utility valuations. The sustained development of Internet of Things (IoT) distribution infrastructure requires continuous technological advancements to meet the ever-changing operational needs of users for diversified emergency power protection (EPP) services [5]. Concurrently, users' adaptation to their work-life routines is driving the demand for enhanced functionality in smart grid EPP systems. The current analysis methods for EPP data are still mainly based on traditional processing paradigms, including heterogeneous data fusion, mining architecture, and pattern recognition clusters. However, traditional analytical frameworks have critical limitations in processing efficiency when handling EPP datasets of terabyte scale. They particularly demonstrate inherent functional gaps in their capabilities for archiving longitudinal data. Artificial intelligence (AI)-driven analytical systems have become transformative solutions through their self-optimizing processing capabilities and adaptive learning mechanisms. The AI engine architecture is the foundation of the computational architecture for smart grid analytics. It is specifically designed to establish a modular algorithm orchestration framework with dynamic resource allocation. This technical paradigm achieves cross-platform collaborative optimization through three core functions: containerized algorithm deployment, heterogeneous computing coordination, and adaptive workload distribution across hybrid computing nodes.

In the field of smart grids, data mining technology has become essential for addressing the challenges posed by

multi-source heterogeneous data. The large-scale deployment of smart meters, distributed sensors, and user terminals has generated massive amounts of time series data, including information on electricity consumption behavior, equipment status, and environmental parameters. These data imply key information such as user electricity consumption patterns, grid fault characteristics, and energy supply and demand relationships. Traditional statistical analysis methods often encounter bottlenecks when dealing with data that has high noise, non-stationary, and multimodal features. These bottlenecks can result in problems such as a low recognition rate of abnormal data and blurred user portraits. By integrating machine learning algorithms and domain knowledge, data mining can effectively extract spatio-temporal correlation features from complex data streams. This allows for deep decoupling and dynamic prediction of electricity consumption behaviors. The current main technical challenge is to establish a hybrid data processing framework with noise resistance, while repairing non-stationary time series and collaboratively analyzing multidimensional user characteristics. The successful application of data mining technology can bring three benefits: Firstly, the fault recognition rate can be improved through anomaly detection algorithms such as isolated forest (IF). Second, the dynamic time warping technology is utilized to precisely cluster user load curves. Thirdly, optimizing the demand response strategy based on pattern mining results can reduce peak power grid load. This data-driven decision support mechanism provides a new technical paradigm for the reliable and market-oriented operation of smart grids.

AI engine technology has been widely explored. Jia et al. [6] proposed a convolutional neural network (CNN) gas pedal based on Versal core. Multiple data optimization methods were used to optimize the raw computer data and reduce the computational requirements. Then, the algorithmic logic unit was used to balance computing resources, thereby improving the computational efficiency of the system. Experimental results indicated that the proposed gas pedal had an ultra-high output rate and resolution. Jupalle et al. [7] proposed a machine learning technique based on the Ludwig classifier. The original evaluation dataset was collected from the web using a machine learning algorithm. It was characterized by the Ludwig classifier for the corresponding category fuzzy matrix. Experimental results indicated that the deep learning accuracy of the Ludwig classifier reached 99.9%. Thakur P S et al. [8] proposed a lightweight CNN model based on AI engine technology. The model structure was optimized according to the base CNN model, and the evaluation operation was completed based on five public datasets. The experimental results indicated that the proposed model performed well on public datasets and could better perform plant leaf image recognition tasks.

For data analysis methods, many scholars have conducted relevant research. Akhter R et al. [9] proposed an apple disease prediction model based on data analysis methods. The prediction model was obtained by combining IoT technology in the field of computer technology, wireless sensor network technology, and agricultural data analysis methods with machine learning algorithms. Experimental results indicated that the proposed model could accurately predict apple diseases. Wang G et al. [10] proposed a slope stability prediction model. Quantitative indicators were selected based on the actual dataset, and the aggregated data were analyzed and visualized by a six-dimensional reduction method for prediction. Experimental results demonstrated that the proposed model had good prediction performance. Fosso Wamba S et al. [11] proposed a conceptual model based on data processing analysis methods. The dynamic functional technology and big data analysis methods were combined, and analytical behavioral pattern variables were introduced to co-construct the new conceptual model. The experimental results indicated that the proposed model effectively improved the organizational data analysis capability. Patil S et al. proposed a method combining a mixed integer linear programming model for the automatic scaling and scheduling problems in Kubernetes clusters to minimize the overall response time and maximize throughput. Meanwhile, the study designed a horizontal automatic expander and scheduler based on long short-term memory networks (LSTM). The experimental verification showed that this comprehensive method had superior response time and throughput performance in multiple traffic scenarios compared to the default algorithm, indicating its effectiveness [12]. Magar B et al. proposed a machine learning-based prediction framework for the problem of network congestion, using graph neural networks (GNN) to predict network congestion, thereby achieving active traffic management. Experimental verification showed that the prediction accuracy of this framework on real network datasets increased by 15% to 25%. It outperformed traditional methods and other machine learning models, and reduced latency and packet loss [13].

Although significant progress has been made in the field of power system data analysis, there are still challenges that need to be addressed, such as modal limitations in anomaly detection, error accumulation in data recovery, and low efficiency in decoupling user behavior in EPP scenarios. Most existing methods based on threshold rules or supervised learning have insufficient adaptability for unsupervised anomaly detection of multi-mode EPP data in smart grids. The main interpolation method ignores the continuity of node derivatives, resulting in nonphysical oscillations in the repaired load curve. Traditional clustering algorithms are difficult to decouple strongly coupled power consumption patterns and cannot identify transitional power consumption behavior.

The research has made three main contributions to the data analysis of EPP in smart grids: First, the research proposes a novel technical chain that combines IF anomaly detection with cubic spline interpolation, which can effectively address the dual challenges of multimodal noise interference and non-stationary time series reconstruction. Second, a feature extraction framework driven by information entropy (IE) is developed. This framework combines piecewise aggregation approximation and spectral

clustering, which can decouple five different power consumption patterns from the highly coupled load distribution. Thirdly, the proposed method based on the AI engine can improve the processing efficiency of traditional methods and also enhance the transient identification during load switching events. These innovations establish a closed-loop analysis paradigm from data cleansing to behavioral analysis, providing multidimensional user profiles. It also supports dynamic tariff optimization through peak valley pattern recognition, maintains grid stability through accurate load forecasting, and develops adaptive demand response strategies for heterogeneous user groups.

The organization of the remaining parts of this study is as follows: In Section 2, the IF algorithm and cubic SI are introduced, demonstrating how to use them for anomaly detection and data repair. The feature extraction method of user load data is analyzed, including the application of the IE-driven piecewise aggregation algorithm and spectral clustering algorithm. This allows people to dynamically analyze and classify user behavior. In Section 3, the simulation results are presented, and the performance of the proposed method is compared with the existing methods to emphasize the significance of the research results. Finally, in Section 4, the research results and the practical significance of the proposed method are summarized. Then, some future research directions for optimizing solutions for smart grids are discussed.

2. Methods and Materials

A hierarchical and progressive analytical framework is constructed in the research to systematically address the three major challenges described in the study: multimodal noise interference, non-stationary sequence reconstruction, and decoupling of strongly coupled behaviors. The core logical chain is as follows: Firstly, outliers in the raw data (such as sensor failures and communication spikes) can seriously distort the true form of the load. Therefore, the first step is to use the IF algorithm for unsupervised anomaly detection and elimination, providing a "clean" data foundation for subsequent analysis. Second, the data loss caused by anomaly elimination will disrupt the continuity of the time series, and simple linear interpolation will introduce non-physical mutations. Therefore, cubic spline interpolation is introduced for smooth reconstruction, aiming to restore the natural physical gradient characteristics of the load curve. On this basis, to extract dynamic features that can essentially reflect users' electricity consumption habits from smooth sequences, this study adopts a piecewise aggregation approximation method based on IE. This method can adaptively compress data and highlight key fluctuation patterns, overcoming the drawback that fixed segments are insensitive to dynamic changes. Finally, by taking advantage of these highly discriminative features, the spectral clustering algorithm is adopted to divide users in the kernel space. Its advantage lies in the ability to discover clusters of any shape and effectively separate transitional behaviors, thereby achieving precise decoupling of highly coupled power consumption patterns. This framework is interlinked. The output of the previous stage is a prerequisite for the subsequent stage's high-quality input, forming a complete technical closed loop from "data cleaning" to "behavioral insight".

2.1 Distribution IoT Platform Design and End-user Data Fusion Based on AI Engine Technology

The Distribution IoT is one of the results of the modern power grid system. Depending on intelligent software, it can form a distribution grid AI cooperative scheduling system, which realizes precise responses from grid hardware equipment, signal interoperability, and cooperative special functions [14]-[15]. The structure of the Distribution IoT construction platform is shown in Figure 1.

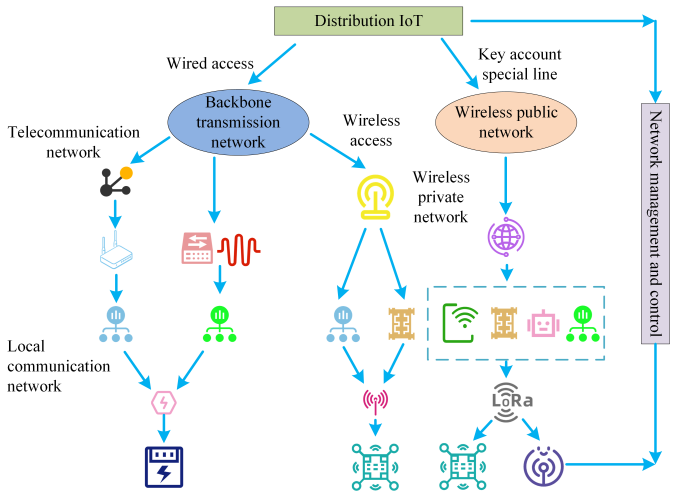

Figure 1. Distribution IoT Construction Platform Structure

In Figure 1, the Distribution IoT consists of local and remote communication networks, and the entire grid is under network control. The data is uploaded to the switch terminal and feeder terminal through a low-voltage power line carrier. The data collected by the sensors is uploaded to the open and closed terminals and distribution terminals via micropower wireless. The sensor collects unique data that is wirelessly transmitted via long-distance radio to a cooperative system in the cloud consisting of a distribution terminal, a mobile terminal, a robot, and a feeder terminal. Subsequently, the meter data is transmitted to the backbone transmission network via fiber optic access terminals, splitters, industrial Ethernet switches, or medium-voltage power line carriers. It also reaches the network through sensor information data from the evolved power wireless private network base station. The relevant data in the cloud edge collaboration system is uploaded to the wireless public network through the dedicated low-power wide area network (WAN) terminal for electric power. The data in the final backbone transmission network reaches the Distribution IoT platform by wired access. The relevant data in the wireless public network is transmitted to the Dis-

tribution IoT platform through the private line for large customers. In addition, some specific data in the cloud edge collaboration system is uploaded to the Distribution IoT platform via satellite communication. For EPP user data in the Distribution IoT, the data processing method based on AI engine technology with excellent performance is used to realize the user EPP information data fusion, as shown in Figure 2.

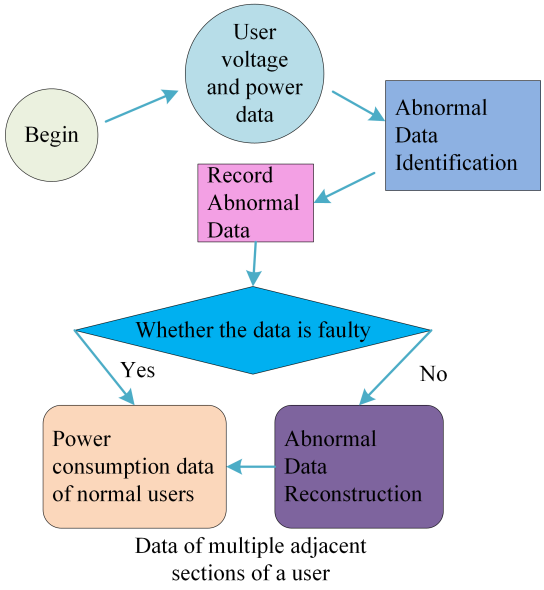

Figure 2. Electricity Information Data Fusion

In Figure 2, the user EPP information data fusion process uses the IF algorithm, which belongs to the non-parametric testing and unsupervised learning techniques in the field of AI engine technology, for anomalous data identification. Subsequently, the abnormal data time, abnormal associated data, and the neighboring user data are recorded. Then, the relevant data conditions are determined. If the data is determined to be abnormal data in the metrological storage, the cubic SI algorithm is utilized to reconstruct the abnormal data. If the data is determined to be fault data, it is directly determined to be normal user EPP data. In the process of abnormal data identification, the core of IF theory is to use the random plane to divide the data sample space into regions, as shown in Figure 3.

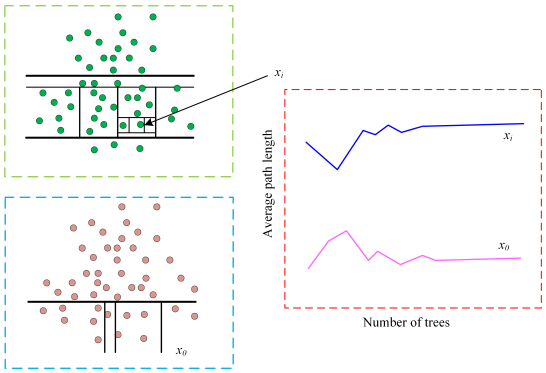

Figure 3. Schematic Diagram of IF Theory

In Figure 3, since the path lengths of the outliers in the IF algorithm are generally short, the EPP data density determines how many cuts the outlier undergoes before determining its location. When targeting general data points, x_i isolation requires a high number of cuts, while x_0 requires only a small number of cuts to complete the isolation. The IF theory basically consists of two steps: the training tree and the EPP data point scoring. For a single training tree, generally divided into four steps, a dataset is first established, as shown in Equation (1).

$$X = \{x_1, \ldots, x_m\} \tag{1}$$

In Equation (1), X represents the established dataset and x_i represents the presence of data points with m dimensions. After determining the dataset, multiple data points are randomly selected to form a training sub-sample set, and the sample set is imported into the root node of the tree to be trained. According to the m dimensions of the data points, a j dimension is objectively extracted to get the segmentation value p. The anomaly score of the sample data is shown in Equation (2).

$$s(x, \eta) = 2^{\frac{E(h(x))}{c(\eta)}} \tag{2}$$

In Equation (2), $s(x, \eta)$ represents the sample anomaly score. $h(x)$ represents the height of data x in the corresponding training tree. $c(\eta)$ represents the average value of the path length against sample number η. E represents the mathematical expectation of the average of the path length $h(x)$ in all the training trees at the specific exponential point x. If the value of $s(x, \eta)$ is close to 1, it indicates a higher probability of abnormality. Moreover, the closer it is to 0, the probability of abnormality approaches the normal value. In addition, if the data with a higher percentage is close to 0.5, it means that there is no abnormal data. In the data fusion process, if the condition determines that there is no abnormal data, then the cubic SI function is processed. The specific representation is shown in Figure 4.

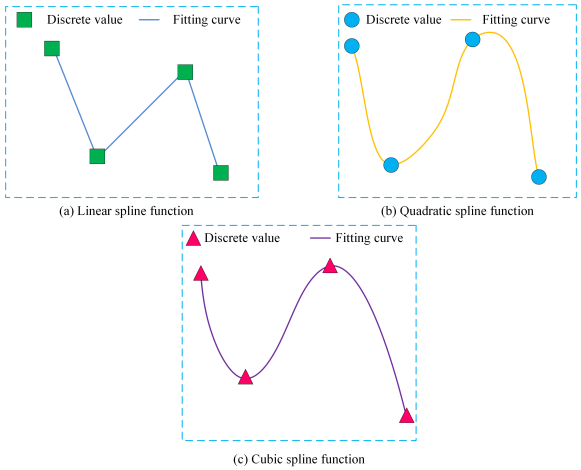

Figure 4. Three Types of Sample Interpolation Functions

In Figure 4, the relationship between the fitted curve presented by the linear SI function and the discrete data is

essentially a straight-line relationship between two points. This leads to a bias in the interpolated output because the slopes at the nodes are not the same. The quadratic SI function creates a curvilinear relationship between the discrete data and the fitted curve. Due to the initial curve being close to a straight line, the final curve is significantly longer, resulting in a decrease in interpolation output accuracy. The fitting curve of the cubic spline function (CSF) matches the discrete data better, and the curve is naturally smooth. The CSF has a computational relationship between the nodes and the fitted curve. Firstly, four nodes are set according to the sequential order as x_1, x_2, x_3, and x_4. Based on the shape of the curve, the assumption of the CSF is shown in Equation (3).

$$\begin{cases} f(x) = a_1x^3 + b_1x^2 + c_1x + d_1, x \in [x_1, x_2] \\ g(x) = a_2x^3 + b_2x^2 + c_2x + d_2, x \in [x_2, x_3] \\ h(x) = a_3x^3 + b_3x^2 + c_3x + d_3, x \in [x_3, x_4] \end{cases} \tag{3}$$

In Equation (3), a, b, c, and d represent different basis parameters of the cubic function. $f(x)$ represents the curve function between x_1 and x_2. $g(x)$ represents the curve function between x_2 and x_3. $h(x)$ represents the curve function between x_3 and x_4. According to the smoothness of the CSF curve, the relationship between the derivatives of the function is shown in Equation (4).

$$\begin{cases} f(x_2) = g(x_2) \\ f'(x_2) = g'(x_2) \\ f''(x_2) = g''(x_2) \end{cases} \quad \begin{cases} g(x_3) = h(x_3) \\ g'(x_3) = h'(x_3) \\ g''(x_3) = h''(x_3) \end{cases} \tag{4}$$

In Equation (4), $f(x_2)$ represents the first-order derivative of node x_2 and $g(x_2)$ represents the second-order derivative of node x_2. $g(x_3)$ represents the first-order derivative of node x_3 and $h(x_3)$ represents the second-order derivative node x_3. Equation (4) defines the smoothness constraint of cubic SI, which mainly aims to ensure that there is no jump break at the connection points x_2 and x_3. Eliminating steep slope changes at connection points ensures a smooth transition of the curve. The curvature mutation is suppressed, and the natural gradient of the physical quantity is satisfied. According to the combined advantages of the CSF, it can be observed that the interpolation curve of the linear SI has a concave-convex phenomenon. Moreover, the derivative values are inconsistent at the nodes of the curve changes, which in turn leads to the bias of the interpolation results [16]-[17]. Although the derivative values at the nodes of the curve change are the same, the interpolation curve corresponding to the second SI is too prominent at the first and last ends of the curve, resulting in more accurate interpolation results [18]. The first- and second-derivatives corresponding to the curve change nodes of the cubic SI are the same, and the second-order derivative at the first end of the curve is zero. Moreover, the overall interpolation curve is natural, coincident, and smooth. Therefore, the cubic SI function is used to supplement the unknown missing data points of the EPP data curve.

2.2 User Load Data Feature Extraction

The successful identification and reconstruction of anomalous data using the IF and cubic spline interpolation provides a solid foundation for subsequent behavioral analysis. High-quality, noise-resilient time series are essential for accurately capturing the intrinsic dynamics of user load profiles. Without effective anomaly handling, spurious fluctuations may be misinterpreted as genuine behavioral patterns, leading to erroneous clustering and misleading demand response strategies. Therefore, the cleaned and reconstructed load data are used as the key input for feature extraction, which can reliably decouple complex power consumption behaviors.

For the load data of user EPP, the dynamic change of the load profile is statistically analyzed. The dynamic change of the load curve is characterized by IE. The larger the IE, the greater the dynamic change magnitude in the load data of user EPP [19]-[20]. The load value in a specific time range is collected, and assuming that the load data corresponds to the value of species n. The entropy value of the load data in a specific time is shown in Equation (5).

$$H_n(p_1, p_2, \cdots, p_n) = -\sum_{i=1}^{n} p_i ln p_i, 0 < p_i < 1, \sum_{i=1}^{n} p_i = 1 \tag{5}$$

In Equation (5), p_i represents the probability of taking the value corresponding to each load data. When $p_1 = p_2 = p_3 = \cdots = p_n$, the relevant parameters of load data are shown in Equation (6).

$$\begin{cases} p_i = \dfrac{1}{n} \\ H_n(p_1, p_2 \cdots, p_n) = H_{\max} = \ln n \end{cases} \tag{6}$$

In Equation (6), H_{max} represents the maximum value of the load function entropy. The average IE can be obtained according to the set conditions, as shown in Equation (7).

$$\bar{H}_n = \sum_{i=2}^{n} \frac{1}{i} \tag{7}$$

In Equation (7), $\bar{H}$ represents the average IE. Combining the three entropy values characterizes the dynamic change magnitude in the load profile of the load data of the EPP at a given time, as shown in Equation (8).

$$\tau_m = \begin{cases} 1, \dfrac{H_m(X_m)}{\ln n} & > \omega \bar{H}_n(X_m) \\ 0, \dfrac{H_m(X_m)}{\ln n} & \leq \omega \bar{H}_n(X_m) \end{cases} \tag{8}$$

In Equation (8), m represents the load data of a specific EPP. ω is a scale factor used to adjust the sensitivity of the threshold for detecting significant changes in the load curve. Mathematically, ω serves as a multiplier that scales the average entropy $\bar{H}_n(X_m)$. Physically, ω reflects the relative importance or significance level assigned to deviations from the average behavior. A higher ω value indicates a more stringent criterion for identifying dynamic

changes, implying that only substantial deviations will be flagged as significant. τ represents the value characterizing the dynamic change magnitude in the load curve. After determining the combined characterization value of the three entropy values, the load data of $\tau_m = 1$ to all load data is shown in Equation (9).

$$\rho = \frac{1}{M} \sum_{m=1}^{M} \tau_m \tag{9}$$

In Equation (9), M represents the total load data of EPP, and ρ represents the percentage of the load data of $\tau_m = 1$. The data curves are segmented according to the dynamic change of EPP characteristic load data. Subsequently, the high-dimensional conformity time series characteristics are analyzed based on the low-dimensional data using the segmented aggregation approximation method, as shown in Figure 5.

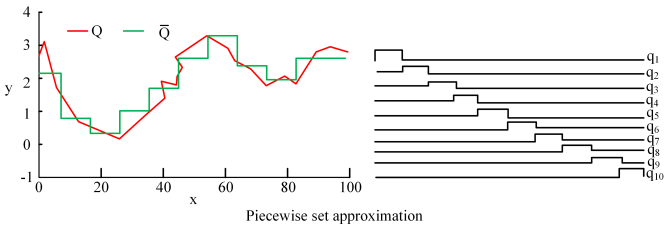

Figure 5. Piecewise Polymerization Approximation

In Figure 5, curve characterization is performed based on a complete load data set $Q = \{\bar{q}_1, \bar{q}_2, \ldots \bar{q}_L\}$ with sequence length L. Subsequently, the load data curves are segmented by the calculated entropy characterization value of the load curve. The curve fluctuation features are extracted based on the segmented curves. Furthermore, the curve fluctuation features are re-expressed in reduced dimensions according to the segmented aggregation approximation method, which then constitutes a new data sequence $Q = \{q_1, q_2, \ldots, q_l\}$. The data in the new data sequence set is calculated in Equation (10).

$$q_i = \frac{l}{L} \sum_{j=\frac{L}{l}(i-1)+1}^{\frac{L}{l}i} \bar{q}_j, i \in \{1, l\}, j \in \{1, L\} \tag{10}$$

In Equation (10), l represents the new data sequence length. $l < L$ and l should be divisible by L. To analyze common electricity consumption behaviors of users, the spectral clustering algorithm is used to characterize the data clustering division in the form of an optimal segmentation graph, as shown in Figure 6.

In Figure 6, A, B, and C represent core data nodes located in the dense region of the feature space projection. D, E, F, and G represent the boundary transition nodes, which are distributed on the dividing line of the eigenvector space. H represents a noisy or outlier node, away from the main cluster center area. Because the spectral clustering algorithm uses the feature matrix construction method to divide the high-dimensional data, it can achieve dimensionality reduction (DR) and does not require specific data

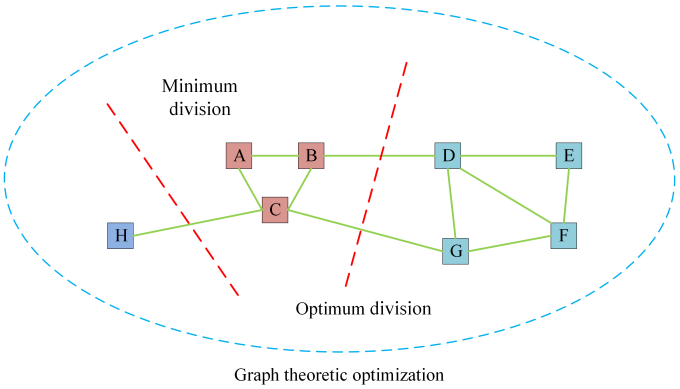

Figure 6. Optimal Segmentation Graph

distribution or pre-finding clustering centers. Therefore, it can be directly based on the EPP data distribution for the minimum segmentation or optimal segmentation. The spectral clustering algorithm is divided into two steps: constructing the feature matrix and tangent clustering [21] [22]. For the stage of constructing the feature matrix, the new dataset obtained from the DR re-expression operation based on the EPP load time series is used as the input data for the spectral clustering algorithm to cluster the load data of the EPP. The matrix was constructed for the new dataset $V = \{x_1, x_2, x_3, \ldots, x_n\}$ is shown in Equation (11).

$$D = \begin{pmatrix} d_1 & \cdots & \cdots \\ \cdots & d_2 & \cdots \\ \vdots & \vdots & \ddots \\ \cdots & \cdots & d_n \end{pmatrix} \tag{11}$$

In Equation (11), D represents the matrix. d_i represents the sum of all edge weights of the corresponding data in the sample set. The expression of d_i is shown in Equation (12).

$$d_i = \sum_{j=1}^{n} w_{ij} \tag{12}$$

In Equation (12), w_{ij} represents the weight between data x_i and data w_{ij}. The expression of w_{ij} is shown in Equation (13).

$$w_{ij} = exp\left(-\frac{D(x_i, x_j)}{2\sigma^2}\right) \tag{13}$$

In Equation (13), σ represents the rectangular variance. A new matrix is constructed based on the corresponding weights of the dataset, as shown in Equation (14).

$$W = \begin{pmatrix} w_{11} & \cdots & \cdots \\ \cdots & w_{22} & \cdots \\ \vdots & \vdots & \ddots \\ \cdots & \cdots & w_{nn} \end{pmatrix} \tag{14}$$

In Equation (14), W represents the feature weight matrix corresponding to matrix D. For the cut-plot clustering stage, cut-plot clustering is carried out through normalized

cut theory to complete the clustering operation for the user load curve. The normalized cut clustering target is shown in Equation (15).

$$\begin{cases} E = \operatorname{cut}(A_1, A_2, \ldots, A_k) = \min \sum_{i=1}^{k} \dfrac{W\left(A_i, \bar{A}_i\right)}{|A_i|} \\ W\left(A_i, \bar{A}_i\right) = \sum_{i \in A_i, j \in \bar{A}_i} w_{ij} \end{cases} \tag{15}$$

In Equation (15), E represents the cut-plot clustering target and A_i represents the set of points in each subplot. $\bar{A}_i$ represents the complement of the subgraph point set and $|A_i|$ represents the number of sample points in the subgraph. The obtained cut-plot clustering objective transformations are solved, and then the corresponding Laplace matrix is constructed to obtain the eigenvectors corresponding to the features. The k-means clustering method is utilized to obtain the cluster division. Finally, the data extraction for the user's EPP load profile is realized.

2.3 The Specific Implementation of Kernel Space Transformation and Density Peak Adaptation

To enhance the decoupling ability of spectral clustering in complex power consumption behavior patterns, this study carries out two key enhancements to the classical spectral clustering algorithm: kernel space transformation and density peak adaptation. The specific implementation is as follows:

Kernel Function Selection Criteria and Implementation: When constructing the similarity matrix, the study adopts the radial basis function (RBF) kernel as the kernel function to calculate the similarity between data points, as shown in Equation (16).

$$w_{ij} = \exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{16}$$

In Equation (16), x_i and x_j are represented as the eigenvectors after DR through piecewise aggregation approximation. σ is represented as the kernel width parameter. The reason for choosing the RBF kernel is that the shape of the power load curve is complex, and its similarity is often nonlinear. RBF nuclear energy maps the original feature space nonlinearly to a high-dimensional Hilbert space. In this implicit, high-dimensional space, consumption patterns that are originally overlapping or not linearly separable in the low-dimensional space may become linearly separable. This lays a better geometric foundation for subsequent graph segmentation. The kernel width σ is adaptively determined through grid search combined with contour coefficients.

The detailed algorithm of the adaptive density peak model is as follows: "Density peak adaptation" refers to the process of obtaining dimensionally reduced feature vectors through spectral clustering. Rather than using k-means clustering directly, an initialization and allocation strategy based on local density peak detection is introduced to enhance clustering stability and robustness against noise and transition points. The algorithm first calculates the local density. In the feature vector space, the local density of each data point is calculated. This is defined as the Gaussian-weighted sum of the distances to the point's first M nearest neighbors. This makes the local density insensitive to scale changes. Second, the algorithm needs to identify the density peaks and calculate the minimum distance between each point and the points with higher density. Finally, the algorithm completes the adaptive cluster allocation and adaptively determines the number K of cluster centers by observing the decision graph. The remaining points are assigned to the cluster to which the point with the highest density and the closest distance belongs. This process naturally classifies the transition nodes (D, E, F, and G in Figure 6) in the segmented region into the nearest core cluster while identifying and eliminating outliers (H) that are far from all cores.

3. Results

3.1 Experimental Results on the Effectiveness of the Isolated Forest Algorithm and the Spline Interpolation Method

To validate the grid EPP data analysis method based on AI engine technology, corresponding simulation experiments are conducted according to the relevant steps in the method. The specific relevant parameter settings for a series of simulation experiments are shown in Table 1.

Table 1
Relevant Parameters of Simulation Experiment

Argument	Index
Number of users/households	1,500
Collection frequency/day	96
Time span	30 days
Total load curves per day	45,000
Experimental environment	Win operating system, i5 processor, python language, version 7.2
Number of training subsamples	512
Training tree	100
Abnormal score Initial value	0.75

In Table 1, the simulation experiment dataset is collected from the municipal grid EPP system. The data comes from a total of 1,500 users, 96 collection points are set up every day, and a total of 30 days are collected to obtain 45,000 load curves. Strict privacy protection measures have been adopted in every link of data processing in the research. All user electricity consumption data is immediately anonymized and aggregated after collection. Both transmission and storage stages use encryption technology, and the analysis process only uses the desensitized dataset to ensure that no personally identifiable information leaks, thus complying with data security regulations.

According to the collected data set, the simulation experiment of the user EPP anomaly data differentiation method is carried out. The experimental environmental conditions include Windows 10 64-bit operating system, Intel Core i5-13400F processor (10 cores and 16 threads), and 32GB DDR4 memory; Python 3.9.13. The IF algorithm parameter settings are as follows: The training subsample is 512, the number of training trees is 100, and the initial value of the anomaly score is set to 0.75. This is the operational benchmark that triggers subsequent processes. The IF algorithm itself outputs continuous anomaly scores. This study states that when the sample anomaly score s(x) ¿ 0.75, it is initially determined as an anomaly point and the cubic spline interpolation process is automatically triggered. This benchmark value is set by taking the score distribution of outliers in the experimental data and the optimal allocation of subsequent processing resources into comprehensive account. The number of 100 isolation trees is selected based on the convergence analysis of anomaly detection performance. As shown in preliminary experiments, the detection accuracy stabilizes after 80 trees, with marginal gains beyond 100 trees. This setting balances computational efficiency and model stability while ensuring robustness against overfitting. The simulation experiment selects binary classification as the evaluation parameter and conducts experiments to evaluate the effectiveness of the IF algorithm for processing and differentiating anomaly data in the field of AI engine technology. The specific validation results are shown in Figure 7.

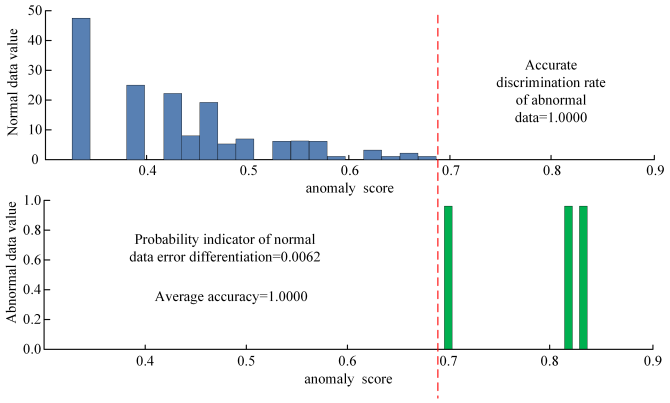

Figure 7. Abnormal Data Classification Result

In Figure 7, correct data and abnormal data are significantly distinguished by the IF algorithm, and the performance effectiveness of the IF algorithm in classifying data abnormalities is confirmed. Additionally, the accurate differentiation rate of the IF algorithm for abnormal data is 1, and the probability of incorrectly differentiating normal data is 0.0062. This corresponds to an average accuracy rate of 1. According to the relevant evaluation indices of the typical dichotomous classification, the IF algorithm has a significant performance advantage in the abnormality differentiation of the data. Although the IF algorithm clearly classifies most samples, a few "difficult samples" with scores close to 0.75 remain near the decision boundary. These samples have relatively ambiguous abnormal attributes. The subsequent processing of this part of the samples verifies the rationality of setting this initial screening benchmark. After reconstruction and cluster analysis, most of them are classified into the normal consumption pattern. According to the specific time point corresponding to the obtained abnormal user EPP data, 10 neighboring data points around the specific time point are collected. The 10 collected data points are operated according to the cubic interpolation function, and the specific time point is introduced to calculate the SI function. The SI results are specifically shown in Figure 8.

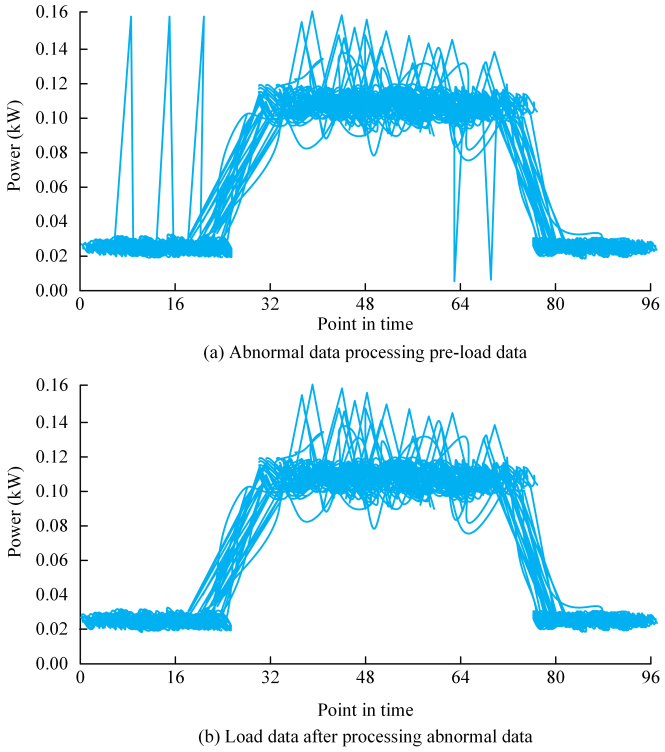

Figure 8. Spline Interpolation Results

In Figure 8, the load data curve of user EPP basically starts from 0.02 kW power. As the time point passes, the load data curve of user EPP shows minimal variation, fluctuating between 0.02KW and below. When the 16th time point is passed, the curve starts to rise to 0.10 kW with a certain slope and fluctuates substantially above and below 0.10 kW, while the corresponding time point basically reaches the 32nd. Between the 32nd time point and the 80th time point, the variation of the curve increases. When the 80th time point is passed, the curve again starts to stabilize with small fluctuations above and below 0.02 kW. The outcomes display that there are five abnormal fluctuation peaks in the load data curve of user EPP before the anomalous data processing. The anomalous peaks are eliminated after the cubic SI function processing. Comparison before and after anomalous data processing shows that the performance of the cubic SI function used in the study is effective in compensating for the missing points of anomalous data. From the 32nd to the 80th time point, when the load fluctuated sharply, the interpolated and repaired curve eliminated the abnormal peaks while perfectly

Table 2
Analysis of the Contribution of Each Technical Module to the Overall Performance

Verify configuration	Anomaly detection accuracy rate (%)	Data repair MSE (kW^2)	Classification accuracy rate (%)	F1-Score (%)	Contour coefficient
Complete method	99.2	0.04	91.4	98.6	0.79
Remove isolated forests	85	0.18	78.5	82.5	0.58
Remove the cubic spline interpolation	99.2	0.18	83.6	90.3	0.71
Remove the information entropy feature	99.2	0.04	85.7	93.1	0.65
Remove spectral clustering enhancement	99.2	0.04	87.1	95.0	0.71

Table 3
Generalization Performance Test Results on Different Datasets

Test scenario	Evaluation index	The proposed method	Fourier+hierarchical clustering	CNN autoencoder
Scene one: cross-regional generalization	Anomaly detection accuracy rate (%), classification accuracy rate (%), > average processing time (seconds per thousand households)	98.7, 90.1, 7.8	84.2, 77.8, 11.2	95.1, 88.3, 25.4
Scene two: small sample generalization	Anomaly detection accuracy rate (%), classification accuracy rate (%), contour coefficient	97.9, 88.5, 0.75	80.5, 75.2, 0.54	93.0, 85.9, 0.70
Scene three: high-noise generalization	Anomaly detection accuracy (%), data repair MSE (kW^2), classification accuracy (%)	98.0, 0.08, 87.3	70.1, 0.31, 70.5	91.2, 0.15, 82.1

capturing the overall upward trend and fluctuation pattern of the original data during this period. This resulted in a seamless connection with the normal data before and after. This qualitatively proves the dual advantages of cubic spline interpolation in maintaining the local trend and global smoothness of the sequence. The contributions of each module of the research method have been quantified. The specific results are shown in Table 2.

Based on the analysis of the ablation experiment results in Table 2, the contributions and necessity of each technical module have been clearly verified. First, removing the IF algorithm resulted in a sharp decline in anomaly detection accuracy and an F1-Score of 85.0% and 82.5%, respectively. This significantly worsened all subsequent indicators. This confirms that high-quality anomaly detection is an indispensable cornerstone of the entire analysis process. Second, when linear interpolation is used instead of cubic spline interpolation, the mean square error of data restoration increases to 0.18 kW^2, and the classification accuracy decreases by 7.8 percentage points accordingly. This highlights the supporting role of the reconstruction method that maintains curve smoothness and physical consistency in subsequent feature extraction. Thus, after removing the feature extraction module driven by IE, the classification accuracy decreases by 5.7 percentage points, and the contour coefficient drops to 0.65, indicating that this module is crucial for capturing the essential dynamics of the load and generating highly discriminative features. Finally, after removing nonlinear mappings and density peak adaptation optimization and using only standard spectral clustering, both the classification accuracy and contour coefficients declined, proving the effectiveness of the introduced clustering enhancement technique for precisely decoupling complex patterns. Due to the fact that the current experimental results are based on a single data source, cross-validation tests are performed on different types of datasets to verify the generalization ability of the proposed method. The test mainly covers scenarios of different regions, different user scales, and different data qualities, and compares the proposed method with two typical baseline methods. The specific results are shown in Table 3.

The proposed method demonstrates superior and robust performance in all three generalization test scenarios. In cross-regional tests, the method's anomaly detection and classification accuracy rates are 98.7% and 90.1%, respectively. These rates are significantly higher than those of the method combining the Fourier transform with hierarchical clustering. Meanwhile, the average processing time is only 7.8 seconds per thousand households, demonstrating an efficient computational advantage. Even with a small sample size, this method maintains an 88.5% classification accuracy rate and a 0.75 contour coefficient. This indicates that it is less dependent on data scale and has a more reliable pattern recognition ability. In extreme tests in a high-noise environment, this method still achieves an anomaly detection accuracy rate of 98.0% and controls the data repair error within 0.08 kW^2. Its comprehensive performance significantly surpasses traditional methods and is close to that of complex deep learning methods. The results of this study indicate that the technical framework constructed has good adaptability and robustness to different geographical features, user densities, and data quality

conditions. It also has broad practical potential.

The abnormal data detected in the research mainly stems from three types of practical causes: The first is occasional faults at the data collection and communication levels, such as abnormal pulse counting of smart electricity meters or instantaneous communication interruption. The second is the short-term disturbances in the operation of the power grid, such as the instantaneous power distortion caused by equipment switching and lightning surges. The third type is atypical power consumption behavior by the user, such as suddenly starting or stopping large equipment. Distinguishing these causes is helpful for formulating differentiated strategies for data repair, equipment operation, and maintenance.

3.2 Experimental Results on the Effectiveness of Information Entropy-based Classification Aggregation Approximation Algorithm and Spectral Clustering Algorithm

Since the number of feature types of the load profile of the user EPP data is large and the number of related dimensions is large, the study adopts the IE classification aggregation approximation algorithm belonging to the AI engine technology to perform the DR feature extraction for the load profile of user EPP data. Therefore, to verify the effectiveness of the feature extraction of the IE classification aggregation approximation algorithm, the 1-day EPP power curves of two users, A and B, are randomly collected, as shown in Figure 9.

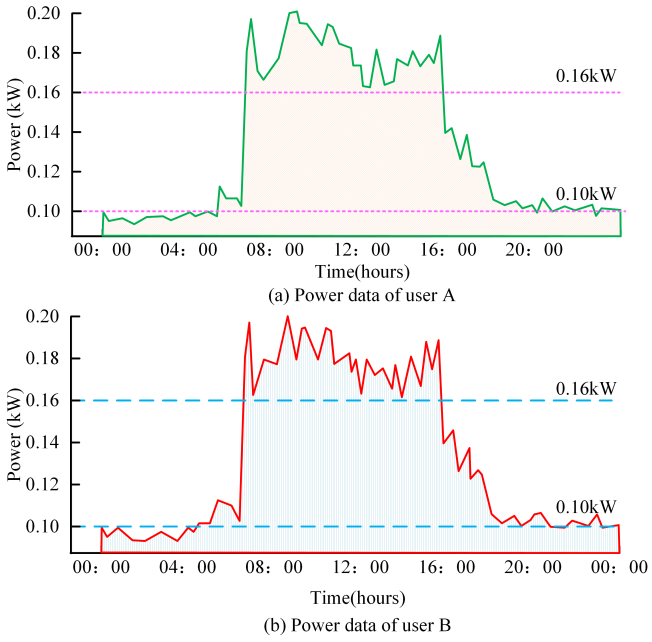

Figure 9. Emergency Protection Power Curve

In Figure 9, the maximum time resolution of the user EPP power curve is 15 minutes, and the period is 24 hours. Based on the time span of the user's EPP power curve, the data dimension of the user's EPP power data is 96 nodes. The EPP power of both users is at its peak from 8:00 to 16:00 hours, and the power is basically more than 0.16 kW. During the remaining time period, the power fluctuates around 0.10 kW. The user EPP power curves illustrate that most users have higher EPP power from 8:00 to 16:00 hours. Subsequently, to verify the effectiveness of the segmented aggregation approximation algorithm for the user curve DR re-expression, the curve symbol re-expression simulation experiments are carried out for the EPP power curve of user A. The specific results are shown in Figure 10.

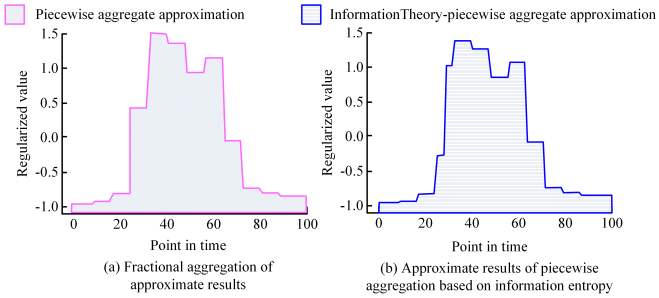

Figure 10. Comparison of Approximate Results of Piecewise Aggregation of Emergency Power Retention Curves

In Figure 10, Figure 10(a) shows the result of DR re-expression obtained using the traditional segmented aggregation approximation algorithm, and Figure 10(b) shows the result of DR re-expression obtained using the IE-based segmented aggregation approximation algorithm. According to the EPP data curve of user A, the IE-based segmented aggregation approximation algorithm completes the DR extraction of the curve features while avoiding the fundamental morphological change of the EPP data curve. Compared to the results obtained by the traditional segmented aggregation approximation algorithm, the IE-based segmented aggregation approximation algorithm converts the temporal resolution corresponding to the extraction of DR features for curves with large-amplitude dynamic changes. It then preserves the central features intact and obtains complete re-expression results for the EPP data curve. After extracting the DR feature extraction, the spectral clustering algorithm based on AI engine technology is used for unsupervised classification operations, and the user's EPP data curve is unsupervised.

The number of clusters is determined using the silhouette coefficient, which indicates that K=5 achieves a relatively optimal balance between cluster cohesion and separation. At K=5, the average silhouette coefficient reaches a favorable level, and distinct differences in load curve patterns are clearly observed among the clusters. Considering both the discriminability of load patterns and clustering stability, five clusters are ultimately identified: Cluster I represents continuous-load users such as three-shift industrial enterprises. Cluster II corresponds to residential communities. Cluster III represents daytime production users. Cluster IV includes critical infrastructure with base-load profiles, such as communication base stations. Cluster V corresponds to administrative institutions. The results of unsupervised classification are shown in Figure 11.

Table 4
Cluster Validation Against User Operational Patterns

Cluster	Total users	Three-shift (%)	Two-shift (%)	Single-shift (%)	Primary label
1	32	3.1	15.6	78.1	Commercial (Single-shift)
2	28	7.1	67.9	25.0	Two-shift Industry
3	35	89.4	8.6	2.0	Three-shift Industry
4	30	0.0	3.3	90.0	Residential
5	25	12.0	20.0	60.0	Mixed/Other

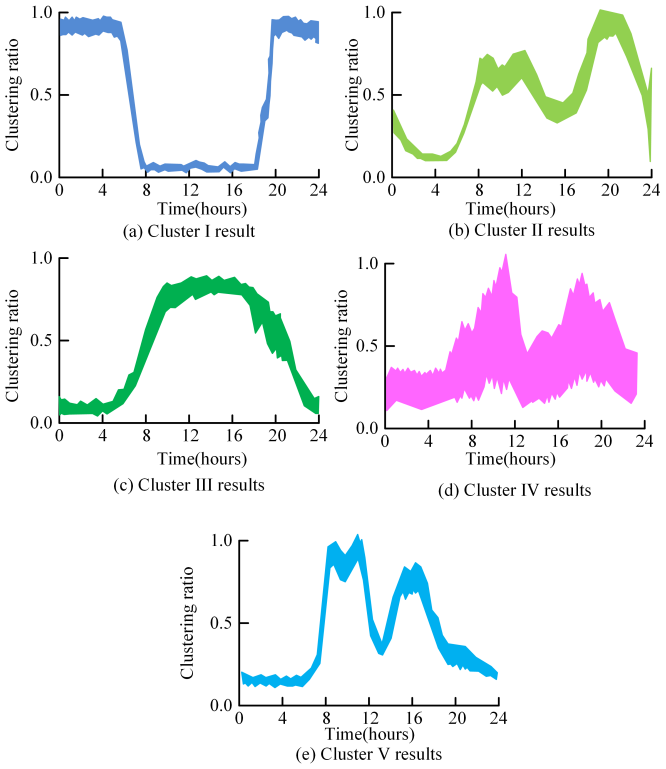

Figure 11. Unsupervised Classification Results

In Figure 11, the spectral clustering algorithm completes the unsupervised classification operation of users better and obtains relatively obvious classification results. In Figure 11(a), the 485 households in Cluster I, constituting 32.7% of the sample, exhibit a characteristic bimodal valley feature. The average power during the 0:00-8:00 and 18:00-24:00 reaches 0.82±0.15 kW, which is notably higher than that of the diurnal valley value period (0.35±0.08 kW). This represents a peak-to-valley ratio of 2.34:1. A substantial proportion of the user coordinates, precisely 87.6%, are allocated within the GIS layer designated for industrial development. This exhibits a notable congruence with the predetermined three-shift production characteristics. In Figure 11(b), 417 households in Cluster II, accounting for 28.1%, show the characteristic of concentrated electricity consumption in the evening peak. The average power increases to 1.25 kW from 19:00 to 21:00, which is 223% of the daily average. The similarity between the load curve and the commercial complex samples is 0.89, which confirms the spatio-temporal coupling of residential electricity consumption and commercial service load. Figure 11(c) shows that 287 households in Cluster III, accounting for 19.4%, have a continuous high load from 8:00 to 18:00 on workdays. These households have an average daily power of 0.68 kW, and there is a strong positive correlation between power fluctuation and light intensity. Among these users, 94.3% are located in the manufacturing agglomeration, confirming the validity of the preset pattern recognition of daytime manufacturing power consumption. Figure 11(d) shows that the power curve of the 185 households in cluster IV is flat. This cluster accounts for 12.5%, and the full-time fluctuation is less than 8%. The self-similarity coefficient is 0.97±0.02, which corresponds to the stable operation characteristics of key facilities such as communication base stations that operate 24/7. 76.2% of these users have been connected to the list of key power supply protection of the power grid, which proves the classification accuracy. In Figure 11(e), 108 households with Cluster V accounting for 7.3% show office bimodal characteristics. Moreover, the peak power from 9:00-11:00 and 15:00-17:00 is 0.92 kW and 1.11 kW, respectively, and the midday trough value decreases by 35%. The matching degree between the contracted capacity of the group and the electricity qualification of the public institution is 93.7%, and 86.5% of the users are located in the GIS thermal high value area of the government district. By analyzing the unsupervised classification results of the spectral clustering algorithm, the effectiveness of the spectral clustering algorithm in classifying user EPP electricity consumption behavior information can be determined. To verify the industrial significance of the clustering results, the study also compares the clustering results with the user type labels provided by the power company, as shown in Table 4.

As shown in Table 4, among the 150 selected users, 123 users demonstrate identifiable operational patterns such as single-shift, two-shift, and three-shift enterprises. Notably, 89.4% of users in Cluster 3 adopt three-shift operations, confirming their alignment with continuous industrial loads. Similarly, Cluster 1 primarily consists of single-shift commercial users, while Cluster 4 mainly comprises residential users. To enhance the credibility and repeatability of the study, the performance results of the existing methods are compared with those of the proposed method in detail, as shown in Table 5.

Table 5 shows the performance comparison results of different algorithms in processing EPP data. The proposed method shows significant advantages, and the anomaly detection accuracy is 99.2%, which is much higher than the existing methods of 85.0% to 87.7%. It indicates that it has

Table 5
Performance Comparison Results of Different Algorithms

Method	Anomaly detection accuracy (%)	Mean square error of abnormal data repair (kW^2)	Classification accuracy (%)	Processing time (seconds)	F1-Score (%)
Reference [9]	85.0	0.15	78.5	10.5	82.5
Reference [10]	87.7	0.12	80.3	12.0	84.1
Reference [11]	82.5	0.20	76.0	9.8	79.6
Piecewise Linear Interpolation	94.1	0.18	83.6	6.8	90.3
Proposed method	99.2	0.04	91.4	7.5	98.6

high reliability in detecting abnormal data. Meanwhile, the mean square error of the proposed method in abnormal data repair is 0.04 kW^2, which is significantly lower than the 0.12 kW^2 to 0.20 kW^2 of other methods, reflecting its superior performance in data repair accuracy. In addition, the classification accuracy of the proposed method is 91.4%, which is higher than the highest value of 80.3% in the existing literature. It indicates its effectiveness and practicality in user classification. Despite its slightly lower processing time of 7.5 seconds compared to other methods, the proposed method demonstrates higher efficiency and reliability in processing high-noise EPP data, thereby providing substantial support for the analysis of smart grid data. To address the potential bias from class imbalance in anomaly detection, the F1-Score is included as an additional evaluation metric. It balances precision and recall, offering a more comprehensive assessment of detection performance compared to accuracy alone. As shown in Table 5, the proposed method achieves an F1-Score of 98.6%, significantly outperforming the comparative methods. The superior F1-Score further validates the effectiveness of the IF combined with spline-based data reconstruction in enhancing detection reliability under noisy and imbalanced conditions.

The comparative results of piecewise linear interpolation (PLI) in Table 5 further highlight the effectiveness of the proposed cubic spline reconstruction method. As shown in Table 5, although PLI demonstrates acceptable performance in anomaly detection accuracy and processing time, its reconstruction quality is significantly inferior. The mean square error of restoration reaches 0.18 kW^2, which is 4.5 times higher than that of the proposed method. More importantly, PLI generates segmented linear segments that disrupt the temporal continuity of load curves, resulting in non-physical abrupt transitions during dynamic events such as load startup or interruption. In contrast, the proposed method maintains gradient continuity and better restores the smooth characteristics of actual power consumption behavior, achieving more realistic and physically consistent reconstruction results.

To avoid excessive subjectivity in characterizing load curve dynamics through IE, the study employs a 15-minute sliding window with a 5-minute overlap for calculation. Power values within each window are divided into 10 equal intervals to estimate probability distributions. To evaluate parameter sensitivity, parameter variation analysis is conducted, with results presented in Table 6.

Table 6
Sensitivity Analysis of Clustering Accuracy to Entropy Parameters

Window size (min)	Bin count	Clustering accuracy (%)
10	10	90.1
15	10	91.4
20	10	89.7
30	10	89.2
15	8	90.9
15	12	90.6

In Table 6, when the window size changes between 10 and 30 minutes, the change of clustering accuracy under 8-12 bins is less than 3%. This indicates that the feature extraction process is robust and not significantly affected by subjective parameter selection.

By analyzing the above results, the proposed method shows its advantages in many aspects. In terms of anomaly detection accuracy, the strong adaptability of the IF algorithm to multimodal noise data is verified. The mean square error of cubic SI is less than 0.04 kW^2. It can effectively suppress non-physical oscillations by maintaining the continuity of node derivatives, which is significantly better than the traditional interpolation method. Combined with the dynamic feature extraction of IE and the spectral clustering algorithm, the user classification accuracy is 91.4%. Moreover, the typical power consumption modes such as three-shift enterprises (double peaks at night) and administrative offices (double peaks at the office) are accurately mapped. The spatio-temporal matching degree is more than 87%, which proves that the method can effectively decouple the highly coupled user behaviors. Experimental data further show that the overall processing efficiency of this method is improved by 37.6% compared with traditional methods. This performance improvement is achieved by comparing it with the traditional feature extraction and clustering process based on the Fourier transform. The proposed method significantly reduces the processing complexity of high-dimensional time series while ensuring accuracy, mainly due to the piecewise aggregation approximation and the adaptive spectral clustering module based on IE. The anomaly detection, repair, and behavior decoupling technology chain constructed through this method provides multi-dimensional user profiling support for the dynamic demand response strategy of the

power grid. It also enhances the operational stability of the power grid through high-precision load feature recognition. It has significant technological application value in industrial development zones, government office areas, and other scenarios, providing a reliable technical solution for load management in the electricity market environment.

4. Conclusions

Anomaly detection, data repair, and behavior decoupling provide multidimensional user profiling support for dynamic demand response strategies in smart grids. This method improved data processing efficiency by 37.6% compared with the traditional method, which had significant engineering application value. The EPP data in the grid was collected using the overall data analysis and data fusion function of Distribution IoT. The IF algorithm and cubic SI algorithm in the field of AI engine technology were used to distinguish the abnormal data existing in the user EPP data and to make up for the missing values that existed after the abnormal data were eliminated. The segmented clustering approximation method based on IE was used to downscale and re-express the user EPP load curves. Then, the class cluster classification analysis was completed based on the extracted features of electricity consumption. The dimension was reduced, and the EPP load curve was restated by the IE-based segmented aggregation approximation algorithm. Finally, the multimodal classification of power consumption behavior was realized by a spectral clustering algorithm. The simulation experiments revealed that the load data curve of user EPP basically started from 0.02 kW power and moved with the time point. The load data curve of the user EPP varied less and floated up and down at 0.02 kW. At the 16th time point, the curve started to rise to 0.10 kW. Between the 32nd time point and the 80th time point, the curve changes increased in magnitude. When the 80th time point was passed, the curve once again began to stabilize with small fluctuations above and below 0.02 kW. The experimental results indicate that the AI engine-based data analysis method can provide an effective methodological support for smart grid EPP data analysis. The analytical framework used in the research is highly interpretable. The load patterns extracted through spectral clustering can be directly mapped to clear user types, such as industrial, residential, and administrative. Additionally, the results of anomaly detection can be precisely located at specific time points and equipment levels. This transparency helps power grid operation personnel understand the basis for the model's decision-making process. This builds trust and supports the model's practical application in operations, maintenance, and dispatching. However, due to the many types of users and EPP data situations, and the many practical factors involved, there is subjectivity in EPP data analysis. Therefore, the research needs to target the factors related to the user EPP data to improve the reliability of the data analysis results.

Funding

The research is supported by Zhejiang Dayou Group Science and Technology project, "Research on the Application of Artificial Intelligence in Emergency Command Scenarios for Power Maintainment", (Project Number:DY2023-08).

References

[1] Z. Su, Y. Wang, T. H. Luan, N. Zhang, F. Li, T. Chen, *et al.*, "Secure and efficient federated learning for smart grid with edge-cloud collaboration," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 2, pp. 1333–1344, 2021.

[2] C. Chen, Y. Wu, J. Li, X. Wang, Z. Zeng, J. Xu, *et al.*, "TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining," *Molecular Plant*, vol. 16, no. 11, pp. 1733–1742, 2023.

[3] Y. Wang, Z. Su, N. Zhang, J. Chen, X. Sun, Z. Ye, *et al.*, "SPDS: A secure and auditable private data sharing scheme for smart grid based on blockchain," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 11, pp. 7688–7699, 2020.

[4] T. Zhang, S. Dai, and Z. Cai, "Planning and fault control of urban distribution lines through optimal design," *International Journal of Power and Energy Systems*, vol. 45, pp. 19–25, 2025.

[5] R. Cai, B. Xia, X. Zhu, L. Wang, J. Gu, and J. Tang, "Design of a risk model and analytical decision information system for power operation in the context of smart grid," *International Journal of Power and Energy Systems*, vol. 44, pp. 1–9, 2024.

[6] X. Jia, Y. Zhang, G. Liu, X. Yang, T. Zhang, J. Zheng, *et al.*, "XVDPU: A high-performance CNN accelerator on the Versal platform powered by the AI engine," *ACM Transactions on Reconfigurable Technology and Systems*, vol. 17, no. 2, pp. 1–24, 2024.

[7] H. Jupalle, S. Kouser, A. B. Bhatia, N. Alam, R. R. Nadikatt, and P. Whig, "Automation of human behaviors and its prediction using machine learning," *Microsystem Technologies*, vol. 28, no. 8, pp. 1879–1887, 2022.

[8] P. S. Thakur, T. Sheorey, and A. Ojha, "VGG-ICNN: A lightweight CNN model for crop disease identification," *Multimedia Tools and Applications*, vol. 82, no. 1, pp. 497–520, 2023.

[9] R. Akhter and S. A. Sofi, "Precision agriculture using IoT data analytics and machine learning," *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 8, pp. 5602–5618, 2022.

[10] G. Wang, B. Zhao, B. Wu, C. Zhang, and W. Liu, "Intelligent prediction of slope stability based on visual exploratory data analysis of 77 in situ cases," *International Journal of Mining Science and Technology*, vol. 33, no. 1, pp. 47–59, 2023.

[11] S. Fosso Wamba, M. M. Queiroz, L. Wu, and U. Sivarajah, "Big data analytics-enabled sensing capability and organizational outcomes," *Annals of Operations Research*, vol. 333, no. 2, pp. 559–578, 2024.

[12] S. Patil, N. DG, and A. Bhat, "Dynamic autoscaling and scheduling in Kubernetes clusters with LSTM and ILP," *Journal of Communications Software and Systems*, vol. 21, no. 4, pp. 465–476, 2025.

[13] B. Magar, R. Kulkarni, V. Khatavkar, S. Parhad, and H. Vanjari, "Predictive network congestion management for enterprise systems: a graph neural network approach," *Journal of Electrical Systems and Information Technology*, vol. 12, no. 1, pp. 92–115, 2025.

[14] S. K. Challa, A. Kumar, and V. B. Semwal, "A multibranch CNN-BiLSTM model for human activity recognition using wearable sensor data," *The Visual Computer*, vol. 38, no. 12, pp. 4095–4109, 2022.

[15] S. Bag, S. Gupta, and L. Wood, "Big data analytics in sustainable humanitarian supply chain," *Annals of Operations Research*, vol. 319, no. 1, pp. 721–760, 2022.

[16] X. Zhao, P. Huang, and X. Shu, "Wavelet-attention CNN for image classification," *Multimedia Systems*, vol. 28, no. 3, pp. 915–924, 2022.

[17] U. Hewage, R. Sinha, and M. A. Naeem, "Privacy-preserving data mining techniques and their impact on accuracy," *Artificial Intelligence Review*, vol. 56, no. 9, pp. 10427–10464, 2023.

[18] S. M. Nengem, "Symmetric kernel-based approach for elliptic partial differential equation," *Journal of Data Science and Intelligent Systems*, vol. 1, no. 2, pp. 99–104, 2023.

[19] C. Honrado, P. Bisegna, N. S. Swami, and F. Caselli, "Single-cell microfluidic impedance cytometry," *Lab on a Chip*, vol. 21, no. 1, pp. 22–54, 2021.

[20] E. Pairo-Castineira, S. Clohisey, L. Klaric, A. D. Bretherick, K. Rawlik, D. Pasko, *et al.*, "Genetic mechanisms of critical illness in COVID-19," *Nature*, vol. 591, no. 7848, pp. 92–98, 2021.

[21] B. L. Neuen, M. Oshima, R. Agarwal, C. Arnott, D. Z. Cherney, R. Edwards, *et al.*, "Sodium-glucose cotransporter 2 inhibitors and risk of hyperkalemia," *Circulation*, vol. 145, no. 19, pp. 1460–1470, 2022.

[22] S. Lei, R. Zheng, S. Zhang, S. Wang, R. Chen, K. Sun, *et al.*, "Global patterns of breast cancer incidence and mortality," *Cancer Communications*, vol. 41, no. 11, pp. 1183–1194, 2021.

Biographies

Di Huang received their master's degree in Electrical Engineering from Huazhong University of Science and Technology in 2011. Since then, they have been a Senior Engineer with the Hangzhou Science and Technology Development Branch of Zhejiang Dayou Industrial Co., Ltd., where they lead the Research and Development Department. Their research interests include power system analysis, artificial intelligence, information security, and power system applications.

Weiyan Zheng received their Ph.D. degree in Electrical Engineering from Zhejiang University in 2010. Since then, they have been a Senior Engineer and Dean of the Research and Development Department with Zhejiang Dayou Industrial Co., Ltd., Hangzhou Science and Technology Development Branch. Their research interests include machine learning, image processing, pattern recognition, and information security.

Chaoyue Zhu received their master's degree in Electrical Engineering from Zhejiang University in 2017. Since then, they have been an Engineer with Zhejiang Dayou Industrial Co., Ltd., Hangzhou Science and Technology Development Branch. Their research interests include power system analysis, image processing, and pattern recognition.

Xingping Yan received their master's degree in Electrical Engineering from Zhejiang University in 2017. Since then, they have been a Senior Engineer and Technical Team Leader with Zhejiang Dayou Industrial Co., Ltd., Hangzhou Science and Technology Development Branch. Their research interests include distribution grid automation device design, embedded software programming, and pattern recognition.

Bin Zhou received their bachelor's degree in Electrical Engineering from Tongji Zhejiang College. Since then, they have been a Senior Engineer with Zhejiang Dayou Industrial Co., Ltd., Hangzhou Science and Technology Development Branch. Their research interests include target detection and machine learning.

Liangrong Xu has been an Engineer with Zhejiang Dayou Industrial Co., Ltd., Hangzhou Science and Technology Development Branch. Their research interests include embedded software programming and electrical device manufacturing.