A Data-driven and Deep Learning-based Method for Power System Operating Mode Identification

Chuan He, *Senior Member, IEEE*, Honghao Zhang, *Student Member, IEEE*, Nan Lu, *Member, IEEE*, Jiawei Hu, *Student Member, IEEE*, Tianqi Liu, *Senior Member, IEEE*, Biao Wang, Jian Gao, and Xi Ye

***Abstract*—Efficient and accurate analysis for power system operating mode identification is crucial for handling the operation, planning, and stability analysis of future power systems. This paper proposes a data-driven and deep learning-based method for power system operating mode identification, while also addressing the needs for identifying future operating modes. Firstly, an operating mode analysis method is developed based on adaptive threshold affinity propagation (AP) clustering with dynamic time warping (DTW), which incorporates historical load sequences and employs a modified distance function and adaptive thresholds for clustering. Secondly, a parallel temporal fusion network (PTFN) model with temporal feature projection is proposed to address load uncertainty. Finally, based on the results of historical operating mode analysis and future load forecasting, a Shapley additive explanation with parallel temporal convolution network and the squeeze-excitation mechanism (SHAP-PTCN-SE) is proposed for identifying future operating modes. Numerical examples demonstrate the efficiency and accuracy of the proposed data-driven and deep learning-based method in identifying future operating modes, providing guidance for system monitoring and protection.**

***Index Terms*—Power system, operating mode identification, parallel temporal fusion network (PTFN), parallel temporal convolution network (PTCN), affinity propagation (AP), data-driven, deep learning, load forecasting.**

I. Introduction

POWER system operating modes refer to a set of identifiable operating modes determined by multi-timescale system load levels under specific system topologies and generator outputs [1]. Power system operating modes, which characterize system states across multiple temporal scales (e.g., yearly, monthly, daily) and resolutions (e.g., hour-level, minute-level, second-level) [2], are primarily influenced by load fluctuations and seasonal variations in generation output, which form relatively stable patterns known as typical operating modes [3]. These modes serve as the foundation for calculating security margins, conducting reliability assessments, and making planning decisions in power systems [4]. Typical examples include various high-peak and low-valley load operating modes affected by seasonal changes and electricity demand of users. The selection of typical operating modes is generally based on seasonal variations and load-level classifications [5], which has been widely validated in practical engineering applications [6]. However, most traditional methods are constrained by expert knowledge and historical modes, which restrict their ability to achieve fine-grained identification in practical engineering scenarios [7].

The rapid development of renewable energy has accelerated the implementation of low-carbon smart grids [8]. With increasing renewable energy penetration, modern power systems exhibit two distinctive features: bidirectional power flows and complex stability mechanisms [9]. These emerging characteristics introduce significant challenges to traditional power system analysis methods, which primarily rely on model-based methods and empirical data [10]. The challenges significantly impact the operation, protection, and stabilization of the power system. Accurate identification of future operating modes directly influences both the security of system operation and the economics of daily dispatch [11]. Errors in distinguishing between peak and valley power system operating modes might lead to misjudgments in reserve capacity and unit commitment, which subsequently increases the dispatch costs and may compromise the capability of responding to contingencies. Therefore, improving the accuracy of identifying future operating modes carries direct significance for both security and economics.

Given that the core of power system operating mode identification involves the analysis of massive datasets, the data-driven method is adopted to capitalize on its technical advantages [12]. Numerous studies have demonstrated the effectiveness of data-driven methods in identifying and analyzing typical operating modes [13]-[15]. In [13], a hybrid method

Manuscript received: July 10, 2025; revised: November 17, 2025; accepted: February 13, 2026. Date of CrossCheck: February 13, 2026. Date of online publication: April 2, 2026.

This work was supported by the Science and Technology Project of State Grid Corporation of China (No. 5108-202218280A-2-263-XG).

C. He, H. Zhang, L. Nan (corresponding author), J. Hu, and T. Liu are with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China (e-mail: he_chuan@scu.edu.cn; zhang_hh@stu.scu.edu.cn; lnan@scu.edu.cn; hjw1126@stu.scu.edu.cn; tqliu@scu.edu.cn).

B. Wang, J. Gao, and X. Ye are with the State Grid Sichuan Electric Power Company, Chengdu 610041, China (e-mail: wangb2850@sc.sgcc.com.cn; gaoj391x@sc.sgcc.com.cn; yex3825@sc.sgcc.com.cn).

DOI: 10.35833/MPCE.2025.000634

combining data-driven methods and interval analysis is proposed to identify bad data in power systems. In [14], a data-driven method is employed to achieve the time pattern discovery, segmentation, and matching of multi-dimensional loads, thereby obtaining typical operating modes. Reference [15] proposes a data-driven method to analyze users' behaviors for constructing a planning decision model in distributed energy systems. Data-driven methods can provide reasonably optimal operational decision schemes for different power systems, regardless of whether they are dominated by conventional units or renewable energy sources [16]. Overall, they have exhibited exceptional performance in addressing identification challenges in power systems, particularly with strong nonlinearity and stochastic behavior [17]. Supervised learning depends on large-scale datasets with accurate sample labels. However, the reliability of historical data gradually declines with changes in power system topology and power flow modes [18]. Furthermore, the hyperparameter sensitivity of unsupervised methods poses another challenge when quantifying similarity between power system operating modes [19]. Besides, as the data dimensionality increases, temporal coupling introduces complex nonlinear correlations and significant redundancy among different dimensions [20]. For example, power flow can be uniquely determined when load and generation data are known. Despite their advantages, data-driven methods often struggle to provide reliable quantitative descriptions of the strong coupling between samples [21]. Consequently, careful selection of appropriate input features is essential to optimize the performance of data-driven methods.

To address the problem of feature redundancy, [22] identifies load level as the primary feature influencing the identification of typical operating modes. However, point clustering methods based solely on daily load levels fail to capture the trend of intra-day load variation, resulting in relatively coarse identification results [23]. Reference [24] minimizes redundancy in load feature dimensions by employing a clustering algorithm combined with a tightness index. However, challenges persist in reducing high-dimensional feature redundancy while considering the trend of intra-day load variation to improve the accuracy of power system operating mode identification and to quantify the temporal distribution features.

Subsequently, previous studies have focused on the identification and analysis of historical power system operating modes, while the analysis of typical operating modes in power systems also requires the development of a reasonable identification strategy for future operating modes. Prior to this, establishing a well-performing load forecasting method can strengthen the overall logical method and provide data support for identifying future operating modes [25]. While data-driven and deep learning methods commonly used for future scenario analysis demonstrate strong identification performance, their black-box feature often obscures the decision-making logic. Power system operators require not only the identification results but also actionable insights into the drivers behind those outcomes [26]. The lack of interpretability makes it difficult to validate the model behavior, build trust from power system operators, and diagnose potential failures [27]. Bridging this gap by developing accurate and interpretable identification models represents a practical necessity for deploying artificial intelligence (AI) solutions in power systems.

Overall, attention is still required to the following three changes. ① How can power system operating modes with high-dimensional nonlinearity be effectively identified? Also, how do targets with varying time durations impact the identification results? ② The future load level of power system exhibit significant uncertainty. How can this problem be addressed to render the proposed model potentially applicable for production in practical projects? ③ How can the accumulation of forecasting errors be mitigated to ensure the accurate identification of future operating modes?

To address the three challenges, a data-driven and deep learning-based method is proposed for the power system operating mode identification, which integrates operating mode analysis, day-ahead load forecasting, and identification of future operating modes. The major contributions of this paper are described as follows.

1) To address the first challenge of power system operating mode identification in high-dimensional nonlinear systems, an operating mode analysis method is proposed based on an adaptive threshold affinity propagation (AP) clustering with dynamic time warping (DTW). Combined with the proposed intra-class evaluation indices, the operating mode analysis method visualizes temporal variations of power system operating modes under different conditions.

2) To address the second challenge of uncertainty in future load level of power system, a parallel temporal fusion network (PTFN) model that considers the projection of temporal features is proposed for day-ahead load forecasting. The PTFN model integrates a long short-term memory (LSTM) projected layer with a parallel temporal convolutional network (TCN), which could better balance both the accuracy and efficiency in day-ahead load forecasting tasks.

3) To address the third challenge of accumulation of forecasting errors in the accurate identification of future operating modes, a Shapley additive explanation with parallel temporal convolution network and the squeeze-excitation mechanism (SHAP-PTCN-SE) is proposed, which innovatively attempts to integrate explainable AI into qualitative characterization of the identification of future operating modes, surpassing traditional methods in both accuracy and interpretability.

The remainder of this paper is organized as follows. Section II presents the framework of the proposed data-driven and deep learning-based method for power system operating mode identification. Section III presents the operating mode analysis method. The PTFN model for day-ahead load forecasting and SHAP-PTCN-SE for identification of future operating modes are presented in Sections IV and V, respectively. Sections VI provides the case study, and Section VII concludes this paper.

II. Framework of Proposed Data-driven and Deep Learning-based Method for Power System Operating Mode Identification

The data of power system operating modes are characterized by high dimensionality, nonlinearity, and strong coupling. To address the three challenges described in Section I, this paper constructs a data-driven and deep learning-based method for power system operating mode identification, and the framework is illustrated in Fig. 1. As shown in Fig. 1, this paper divides the analysis and identification of power system operating modes into three parts.

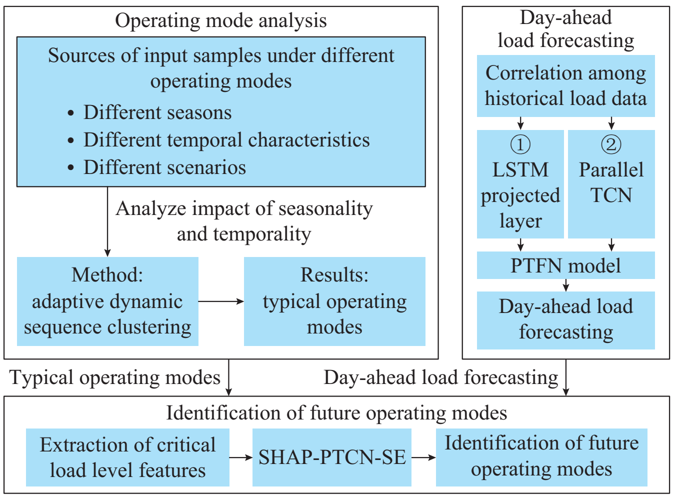

Fig. 1. Framework of proposed data-driven and deep learning-based method for power system operating mode identification.

A. Operating Mode Analysis

An adaptive threshold AP clustering with DTW is developed, which utilizes hour-level historical load sequences as inputs to identify typical operating modes. Additionally, samples with distinct climatic characteristics and varying time lengths are selected to analyze the impact of seasonality and temporality on the classification of typical operating modes.

B. Day-ahead Load Forecasting

By embedding an LSTM projected layer into the parallel TCN, temporal features are essentially extracted through projection, while a parallel framework is employed to enhance the computational efficiency of TCN. Moreover, by considering the correlations among historical load data, a PTFN model for day-ahead load forecasting with temporal feature projection is presented.

C. Identification of Future Operating Modes

With the typical operating modes and the day-ahead load forecasting results, SHAP is employed to extract critical load-level features that have a significant impact on PTCN-SE, which are then used as input to identify future operating modes.

III. Operating Mode Analysis Method

A. Method Description and Algorithm Selection

Practical power system operating modes are relatively complex and diverse, exhibiting strong temporal features. For historical samples with different time lengths, if the identification relies solely on expert knowledge or seasonal factors, the obtained typical operating modes may exhibit significant deviations from actual power system operating modes. Identifying power system operating modes is a typical big-data analytic problem. As unsupervised learning methods, clustering algorithms demonstrate the capability to objectively analyze the differences between different power system operating modes. In contrast to traditional clustering algorithms, the AP clustering algorithm eliminates the need to predefine both the initial cluster quantity and centroid positions, thereby reducing methodological subjectivity. Consequently, it is not necessary to use algorithms such as the Silhouette coefficient to determine the cluster number, resulting in an optimized architecture with enhanced computational efficiency. The operating mode analysis method is illustrated in Fig. 2. The gray curves represent time-series load samples within a cluster, while the red curve corresponds to the cluster centroid.

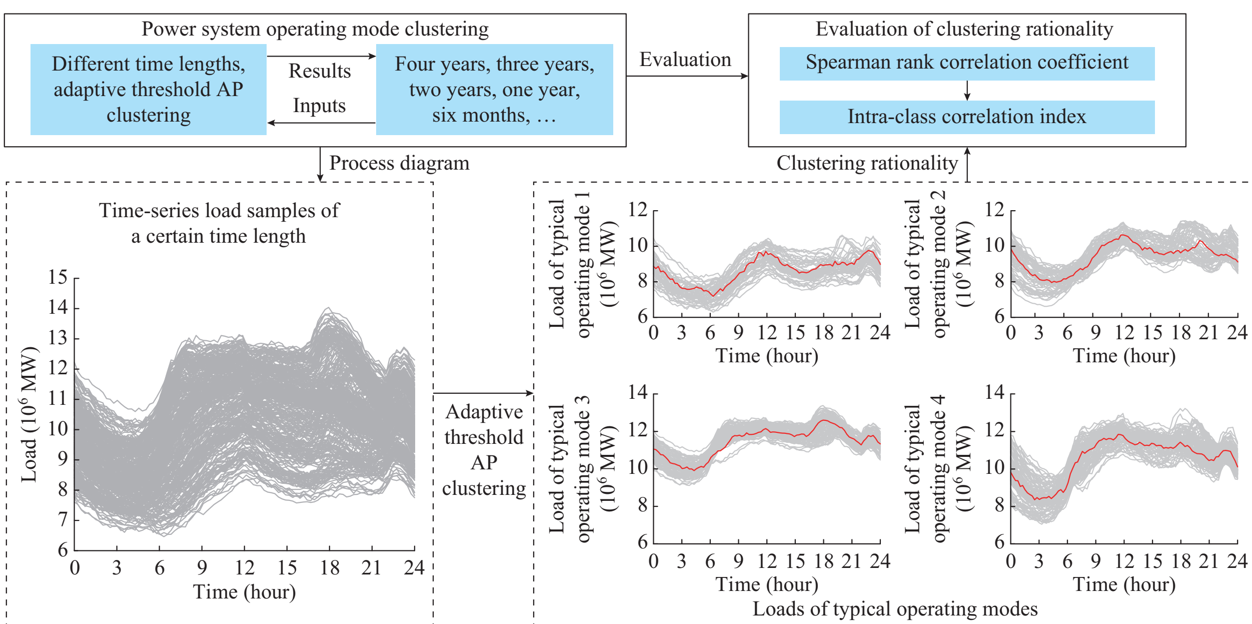

Fig. 2. Illustration of operating mode analysis method.

By taking time series of different lengths as input, the adaptive threshold AP clustering with DTW is employed to analyze the characteristics of power system operating modes. The intra-class correlation index based on Spearman rank correlation coefficient is designed to evaluate the clustering rationality of clustering results, verifying the effectiveness of operating mode analysis method.

B. Adaptive Threshold AP Clustering with DTW

AP clustering does not need to predetermine the number of cluster exemplars (i.e., target clusters) [28]. The core principle of AP clustering is to treat each data point as a potential exemplar. It iteratively determines optimal exemplars based on similarities for clustering. It quantifies the similarity coefficient $s_{i,k}$ between sample points X_i and X_k using negative Euclidean distance in (1), which evaluates the suitability of X_k as the clustering centroid for X_i.

$$s_{i,k}=-\|X_i-X_k\|^2 \tag{1}$$

The responsibility $r_{i,k}$ and availability $a_{i,k}$ represent the suitability of X_k serving as the clustering centroid of X_i from the perspectives of X_i and X_k, respectively. The calculations of $r_{i,k}$ and $a_{i,k}$ are given in (2) and (3), respectively. Both $r_{i,k}$ and $a_{i,k}$ are iteratively updated in (4).

$$r_{i,k}=s_{i,k}-\max_{j\neq k}\{a_{i,j}+s_{i,j}\} \tag{2}$$

$$a_{i,k}=\begin{cases}\min\left\{0,\ r_{k,k}+\sum\limits_{i'\notin\{i,k\}}\max\{0,\ r_{i',k}\}\right\} & i\neq k\\ \sum\limits_{i'\neq k}\max\{0,\ r_{i',k}\} & i=k\end{cases} \tag{3}$$

$$\begin{cases}r_{i,k}^{t+1}=\gamma r_{i,k}^{t}+(1-\gamma)r_{i,k}^{t+1}\\ a_{i,k}^{t+1}=\gamma a_{i,k}^{t}+(1-\gamma)a_{i,k}^{t+1}\end{cases} \tag{4}$$

where γ is the damping coefficient that satisfies $\gamma\in[0,1]$, which regulates the iterative refinement of both $r_{i,k}$ and $a_{i,k}$; and t is the index of iteration.

Whereas conventional methods typically employ a fixed γ, the selection of the optimal γ often relies on empirical knowledge and manual parameter tuning. This significantly limits the flexibility of models. To address this limitation, this paper proposes to adaptively select γ based on grid search while incorporating the Silhouette coefficient.

Furthermore, this paper focuses on clustering the time series rather than sample points of load, whereas Euclidean distance and DTW [29] represent the two typical distance measurement methods for time series analysis. DTW essentially matches data with similar shapes by identifying an optimal warping path P, thereby endowing the algorithm with enhanced robustness and adaptability [30]. The element $P_k=(x,\ y)_k$ of the optimal warping path P can be calculated as:

$$d_{\mathrm{DTW}}(x,y)=\min_{P}\left\{\frac{1}{K}\sum_{k=1}^{K}d_{\mathrm{ED}}(P_k)\right\} \tag{5}$$

$$d_{\mathrm{ED}}(P_k)=\|x-y\|_2 \tag{6}$$

where K is the sequence length; x and y are the load sequences to be processed; $d_{\mathrm{DTW}}(x,y)$ is the DTW distance between $\boldsymbol{x}$ and $\boldsymbol{y}$; and $d_{\mathrm{ED}}(P_k)$ is the Euclidean distance between x and y.

The steps of adaptive threshold AP clustering with DTW are summarized in Algorithm 1.

Algorithm 1: adaptive threshold AP clustering with DTW

Require: dataset $X=\{X_1,X_2,...,X_n\}$, initial damping coefficient γ_0, step size $e\in(0,1)$, the maximum number of iterations $t_{\max}$, and the maximum number of clusters $N_{\max}$

Ensure: optimal number of clusters N_{best} and optimal damping coefficient γ_{best}

1. Set iteration parameters $i\leftarrow 0$, $k\leftarrow k\neq i$, $m\leftarrow 0$
2. **while** the number of clusters at the m^{th} iteration $N_m<N_{\max}$ or the damping coefficient at the m^{th} iteration $\gamma_m<0$ **do**
3. **for** $t=1$ to $t_{\max}$ **do**
4. Compute the similarity coefficient $s_{i,k}$ with DTW, and $s_{i,k}=-d_{\mathrm{DTW}}$ with (5) and (6)
5. Compute $r_{i,k}$, $a_{i,k}$, and exemplar $c_{i,k}\leftarrow r_{i,k}+a_{i,k}$ with (2) and (3)
6. Update r^{t+1} and a^{t+1} with γ_m and (4)
7. Update $c^{t+1}\leftarrow r^{t+1}+a^{t+1}$
8. **end for**
9. Obtain optimal exemplar $c_m^{\text{best}}\leftarrow\arg\max\{c_m^1,c_m^2,\dots,c_m^{t_{\max}}\}$
10. $N_m^{\text{best}}\leftarrow N_m^{\text{best}}=\|c_m^{\text{best}}\|_0$, where $\|\cdot\|_0$ counts the non-zero elements
11. Compute intra-cluster distance $a(m)\leftarrow\dfrac{1}{|c_m^{\text{best}}|}\sum\limits_{\{X_i\in c_m^{\text{best}}\}}d_{\mathrm{DTW}}(X_m,X_i)$, nearest-cluster distance $b(m)\leftarrow\min\limits_{\{c_i\notin c_m^{\text{best}}\}}\left\{\dfrac{1}{|c_i|}\sum\limits_{\{X_i\in c_i\}}d_{\mathrm{DTW}}(X_m,X_i)\right\}$, and Silhouette coefficient $S_c(m)\leftarrow\dfrac{b(m)-a(m)}{\max\{b(m),a(m)\}}$
12. Update damping coefficient: $\gamma_{m+1}\leftarrow\gamma_i-e(m+1)$
13. **end while**
14. Obtain optimal damping coefficient $\gamma_{\text{best}}\leftarrow\arg\max(S_c(m))$
15. Obtain N_{best} and compute $\{\gamma_{\text{best}},r,a,S\}$
16. **Return**: N_{best} and γ_{best}

Adaptive threshold AP clustering with DTW is proven suitable for power system operating mode identification, owing to its clustering accuracy and ability to automatically determine the number of power system operating modes. When addressing such problems, the computational requirements of the algorithm align effectively with the characteristics of annual datasets, which typically comprise hundreds to thousands of daily load profiles, thereby providing a reliable solution for power system operating mode identification.

C. Clustering Rationality as Performance Evaluation Criterion

The intra-class correlation index based on Spearman rank correlation coefficient is designed to evaluate the clustering rationality and correlation of power system operating mode clustering results. Spearman rank correlation coefficient is an evaluation index widely used to describe the correlation between feature vectors, with its value range bounded by $[-1,1]$. A coefficient of 1 or -1 indicates a strictly monotonic functional relationship between two variables, whereas a value of 0 indicates their linear independence. The Spearman rank correlation coefficient γ_s is defined as:

$$\gamma_s=\frac{\sum\limits_{i=0}^{n}(R_i-\overline{R})(S_i-\overline{S})}{\sum\limits_{i=0}^{n}(R_i-\overline{R})^2\sum\limits_{i=0}^{n}(S_i-\overline{S})^2}\qquad \gamma_s\in[-1,1] \tag{7}$$

where n is the number of variables; $\overline{R}$ and $\overline{S}$ are the mean values of variables in sequences R and S, respectively; and R_i and S_i are the ranks of the i^{th} sample point in R and S, respectively.

The probability distribution of the resulting correlation matrix $\boldsymbol{\Gamma}_s$ is computed over all clusters, where $\boldsymbol{\Gamma}_s$ is derived from the Spearman rank correlation coefficients between each sequence in each cluster and their respective centroid. The clustering performance is characterized by calculating five indices: maximum, minimum, mean, median, and standard deviation (SD) of $\boldsymbol{\Gamma}_s$, which are expressed as:

$$\Gamma_s^{\max} = \max(\boldsymbol{\Gamma}_s) \tag{8}$$

$$\Gamma_s^{\min} = \min(\boldsymbol{\Gamma}_s) \tag{9}$$

$$\Gamma_s^{\text{mean}} = f_{\text{mean}}(\boldsymbol{\Gamma}_s) \tag{10}$$

$$\Gamma_s^{\text{median}} = f_{\text{median}}(\boldsymbol{\Gamma}_s) \tag{11}$$

$$\Gamma_s^{\text{SD}} = f_{SD}(\boldsymbol{\Gamma}_s) \tag{12}$$

where $\Gamma_s^{\max}$, $\Gamma_s^{\min}$, Γ_s^{mean}, Γ_s^{median}, and Γ_s^{SD} are the maximum, minimum, mean, median, and SD of $\boldsymbol{\Gamma}_s$, calculated by the functions max, min, f_{mean}, f_{median}, and f_{SD}, respectively.

IV. PTFN Model for Day-ahead Load Forecasting

A. Framework Description and Algorithm Selection

Extensive studies have demonstrated the outstanding performance of deep learning in solving regression problems. Consequently, given appropriate input features, deep learning methods can essentially handle load forecasting problems. To address the randomness and volatility of power loads, this paper proposes the PTFN model incorporating the projection of temporal features for day-ahead load forecasting. The PTFN model employs TCN as its base algorithm, chosen for its efficiency and accuracy demonstrated in identification tasks. It also introduces targeted improvements for the day-ahead load forecasting objectives.

To enhance the efficiency of day-ahead load forecasting, this paper does not adopt a recursive prediction strategy. Based on time-series forecasting principles, optimal historical load data are selected as model inputs through analysis of Spearman rank correlation coefficients between corresponding moments across the preceding N days. The framework of load forecasting is shown in Fig. 3, where the blue box represents the input sequence of historical loads (e.g., P_{T-1}^t is the load at time t for the day prior to the target forecast day T), the yellow box represents the extracted feature (e.g., P_{input}^t is the feature input to the model), and the pink box represents the final forecast (e.g., P_T^t is the load forecast for day T made by the PTFN at time t).

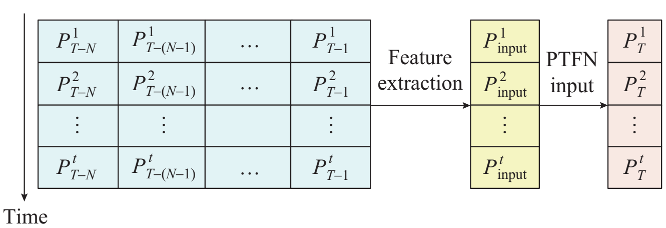

Fig. 3. Framework of load forecasting.

B. PTFN Model

TCN effectively combines the feature extraction ability of convolutional networks with the temporal mining ability of recurrent neural networks. By introducing the dilated causal convolution (DCC) and residual connection (RC) modules, TCN achieves significant reduction in network depth compared with deep convolutional structures such as ResNet and Denseness [31]. Therefore, it can greatly improve the computational efficiency while simplifying the structure.

Each residual block in TCN is characterized by a parameter pair (k_c, d), where $k_c \in \mathbf{Z}^+$ is the convolutional kernel size; and $d \in \mathbf{Z}^+$ is the dilation coefficient. RCs mitigate gradient vanishing and enhance the performance of deep networks [32].

DCC is critical to TCN. Based on causal convolution, DCC improves the dilation coefficient d. Specifically, DCC prioritizes the application of causal convolution, ensuring that historical data are not prematurely accessed, thereby preventing information leakage [31]. Moreover, the network perception layer is extended by increasing the expansion coefficient to enhance the feature extraction ability. The calculation is given by:

$$F = \sum_{v=0}^{u-1} f(v) X_t \tag{13}$$

where F is the dilation convolution operation; X_t is the value of sequence data; $f(v)$ is the filter function; u is the length of the input sequence; and v is the v^{th} element in the input sequence.

The serial structure of TCN leads to the gradual loss of local information when stacking multiple DCC layers, consequently impairing the capability of models to extract relational features from input data [33]. In addition, it is difficult for TCN to acquire both temporal features and depth features, which limits the further improvement of forecasting accuracy [34].

Compared with other deep learning-based methods, the integration of TCN and LSTM networks enables the capability to recognize both short-term and long-term temporal cognitions [35]. Notably, LSTM projected layer possesses a stronger perception capability than LSTM network, but with lower efficiency. However, most of these methods employ an external integrated combination structure, which is structurally simple while exhibiting poor stability. To address these limitations, this paper proposes to integrate LSTM projected layers with TCN through fundamental structural modifications. The resulting PTFN model for day-ahead load forecasting is illustrated in Fig. 4, where $Conv(\cdot)$ represents the convolution function; k_1 and d_1 are the parameters in the DCC; and GELU is short for the Gaussian error linear unit. The specific structural improvements are as follows.

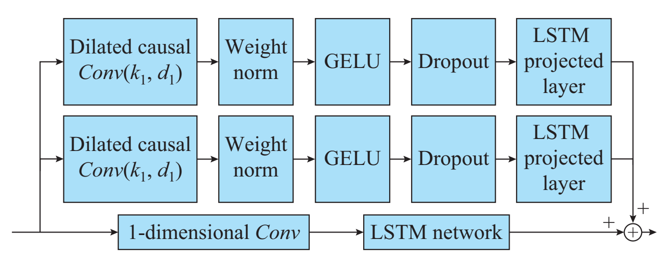

Fig. 4. PTFN model for day-ahead load forecasting.

1) Residual blocks are modified from a serial structure to a parallel structure to avoid information loss caused by excessive layer stacking.

2) LSTM projected layer is inserted after each DCC module to enhance the ability to mine temporal features and improve the perception capability of the PTFN model through projection.

3) LSTM network is inserted after the convolutional layer in RC, ensuring both temporal feature extraction and computational efficiency.

4) The rectified linear unit (ReLU) activation function is replaced with GELU [36] to improve the learning performance.

The details of architecture configuration and hyperparameter selection of the PTFN model can be found in Supplementary Material A Table SAI.

The loss function is chosen as mean square error (MSE) I_{MSE}, which is calculated as:

$$I_{\mathrm{MSE}}=\frac{1}{n}\sum_{i=1}^{n}(y_i-y_i') \tag{14}$$

where y_i and y_i' are the i^{th} true value and forecasting result, respectively.

According to Adam optimization method [37], the prediction error is corrected by designing the adaptive learning rate to control the parameter update rate. Early stopping strategy is used to set termination conditions for model training [38].

C. Forecasting Accuracy as Performance Evaluation Criteria

This subsection selects root mean square error (RMSE) I_{RMSE} and mean absolute percentage error (MAPE) I_{MAPE} to evaluate the performance of regression algorithms. These indices could reflect errors in forecasting results of models from different aspects. Their specific definitions are given as:

$$I_{\mathrm{RMSE}}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i-y_i')} \tag{15}$$

$$I_{\mathrm{MAPE}}=\frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i-y_i'}{y_i}\right| \tag{16}$$

V. SHAP-PTCN-SE for Identification of Future Operating Modes

A. Framework Description and Algorithm Selection

Similar to their outstanding performance in regression problems, deep learning methods, particularly TCN, also perform well in solving identification problems. Building upon the TCN with targeted enhancements, this paper proposes SHAP-PTCN-SE for identifying future operating modes, capable of handling complex identification targets while maintaining reliability.

Sample labels are generated based on the identification results of typical operating modes in Section III. The training set inputs consist of historical load data and critical features such as daily mean load, daily load rate, peak-valley difference, and daily maximum load. Additionally, SHAP-based feature interpretation is used during training to identify the most sensitive features for typical modes, reducing dimensionality and eliminating redundancy. Future load forecasting data obtained in Section IV serve as test set inputs to evaluate the identification performance. The framework for identification of future operating modes is shown in Fig. 5.

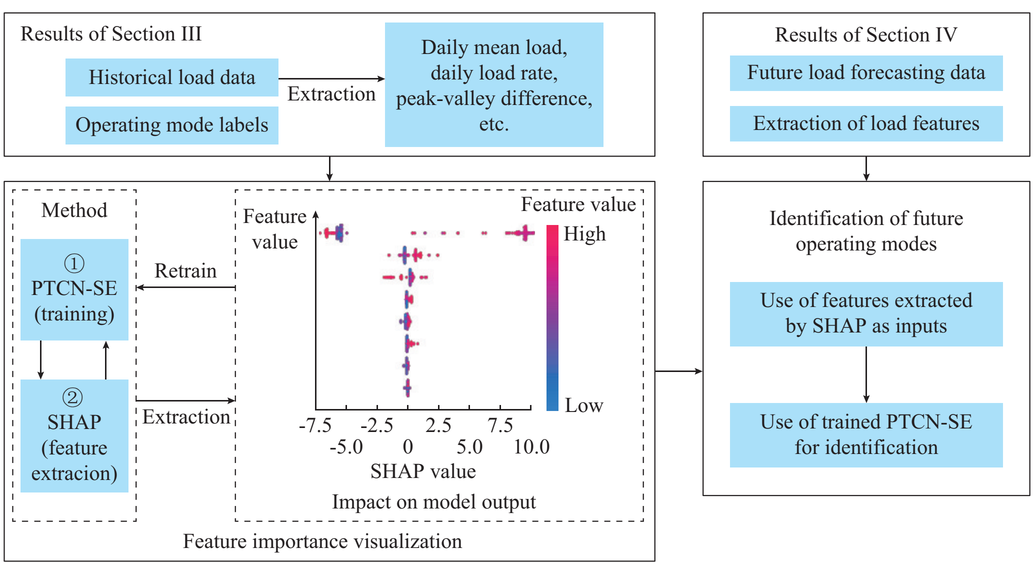

Fig. 5. Framework for identification of future operating modes.

B. PTCN-SE

Compared with regression problems, identification problems place greater emphasis on feature extraction. To ensure the stability of the combined structure whereas considering the strength and limitations of TCN in identification problems, this paper modifies the original RC structure of TCN. SE-net can adaptively recalibrate the importance of channel features to enhance the identification performance of the target network [39]. Based on the above analysis, adopting a similar improvement method as PTFN model, PTCN-SE is used for identifying future operating modes, as shown in Fig. 6. The improvements are described as follows.

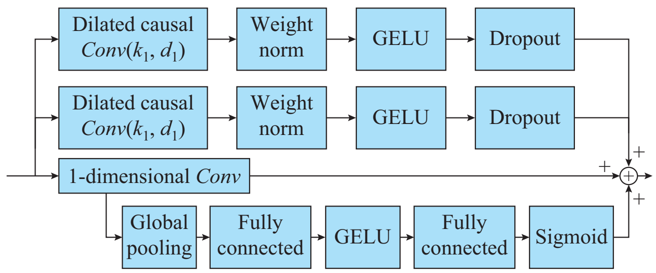

Fig. 6. Illustration of PTCN-SE.

1) Residual blocks are modified from a serial structure to a parallel structure.

2) The SE-net mechanism is introduced into the RC layer, the RC layer is modified into a special SE-net module, and the ability to extract features is improved.

3) The ReLU activation function is replaced by GELU.

The details of architecture configuration and hyperparameter selections of the PTCN-SE can be found in Supplementary Material A Table SAII.

The loss function L is chosen as binary cross-entropy loss (BCE), which is calculated as:

$$L(y_i, y_i') = -\sum_{i=1}^{N^{\text{label}}} [y_i \lg y_i' + (1-y_i)\lg(1-y_i')] \tag{17}$$

where N^{label} is the number of label classes.

To mitigate the effects of an imbalanced training dataset, the weights of the loss function are adjusted to balance the attention of samples [40], thereby enhancing the classification performance of the model on unbalanced datasets. The loss function in (17) is modified as:

$$L(y_i, y_i') = -\sum_{i=1}^{N^{\text{label}}} [\omega_i^+ y_i \lg y_i' + \omega_i^- (1-y_i)\lg(1-y_i')] \tag{18}$$

$$\begin{cases} \omega_i^+ = \dfrac{N_i^{\text{train}}}{N_i^{\text{train}+} N^{\text{label}}} \\ \omega_i^- = \dfrac{N_i^{\text{train}}}{N_i^{\text{train}-} N} \end{cases} \tag{19}$$

where ω_i^+ and ω_i^- are the corresponding weights of the i^{th} positive and negative samples, respectively; N_i^{train} is the total number of samples of the i^{th} label; and $N_i^{\text{train}+}$ and $N_i^{\text{train}-}$ are the numbers of positive and negative samples of the i^{th} label, respectively.

The Adam optimization method is employed for model training, with the early stopping strategy adopted to prevent overfitting.

C. Feature Explanation Based on SHAP

The SHAP is a post-hoc explanation method in machine learning that combines ideas from game theory and local explanation [41]. When explaining the identification results of power system operating modes under any given state, the SHAP performs PTCN-SE locally, constructs a local explanation model, re-identifies the power system operating modes, and produces a calculation result identical to that of the original model. The DeepExplainer, specifically designed for deep learning methods, is employed as the SHAP explainer. The explainer is initialized using the training data of case studies as the background distribution. It calculates the Shapley value for each feature in the local explanation model, which represents the contribution of features to the result under that state. This paper employs SHAP to perform dimensionality reduction on input features, selecting variables with high contribution as inputs for PTCN-SE.

This paper adopts daily mean value, load rate, peak-valley difference, maximum value, minimum value, median, quartile, and third quartile as the initial input features for PTCN-SE. These indices can comprehensively characterize the level, profile shape, and distribution robustness of load. Specifically, indices such as daily mean, maximum, minimum, and median define the anticipated load level of the power system, directly determining the total generation dispatch and main transformer load required under future operating modes. Indices such as peak-valley difference and load rate can reflect significant differences in power system operating modes, making them essential for identifying typical operating modes. Quartile and third quartile can reveal the stability and predictability of the expected load. PTCN-SE based on input features is constructed as:

$$R_{id} = PTCN\text{-}SE(x_{\text{input}}^1, x_{\text{input}}^2, ..., x_{\text{input}}^N) \tag{20}$$

where R_{id} is the identification result of power system operating modes; $PTCN\text{-}SE(\cdot)$ is the constructed PTCN-SE; and x_{input}^i $(i=1,2,\dots,N)$ is an input feature.

In SHAP, all features are considered as contributors, and the contribution of each feature to the output can be represented by the local explanation model. Let $y(x_{\text{input}})$ and $g(x)$ be the functions of the original and local explanation models, respectively. The relationship between them is given as:

$$g(x_{\text{input}}^i) = \phi_0 + \sum_{l=1}^{n} \phi_l = y(x_{\text{input}}^i) \tag{21}$$

where ϕ_0 is the Shapley baseline value; and ϕ_l is the Shapley value of the l^{th} feature.

The Shapley value of the l^{th} feature is calculated as:

$$\phi_l = \sum_{S_{\text{f}} \subseteq N_{\text{f}} \backslash \{l\}} \frac{|S_{\text{f}}|!(N_{\text{f}} - |S_{\text{f}}| - 1)!}{N_{\text{f}}!} (x(S_{\text{f}} \cup \{l\}) - x(S_{\text{f}})) \tag{22}$$

where N_{f} and S_{f} are the numbers of features in full set and subset, respectively; $x(S_{\text{f}} \cup \{l\}) - x(S_{\text{f}})$ is the marginal contribution of l to S_{f}; and $(|S_{\text{f}}|!(N_{\text{f}} - |S_{\text{f}}| - 1)!)/N_{\text{f}}!$ is the weight of the feature. $S_{\text{f}} \subseteq N_{\text{f}} \backslash \{l\}$ expresses the need to sum over all features.

Based on (22), the contribution degree of each feature in each sample is obtained. To evaluate the overall contributions of different features to power system operating modes, the contributions of all samples for the feature are considered, and its mean value is the SHAP value. The larger the SHAP value of a certain feature is, the more important the feature will be. The importance of all features is sorted and selected with a larger overall contribution to the output as inputs for PTCN-SE to reduce the dimensionality.

D. Identification Accuracy as Performance Evaluation Criterion

For visually evaluating the reliability of identification results, the identification accuracy A is introduced as:

$$A=\frac{1}{n}\sum_{i=1}^{\beta}\frac{T_{P,i}}{n} \tag{23}$$

where β is the number of categories; and $T_{P,i}$ is the number of correctly predicted samples for the i^{th} category.

VI. Case Study

Case study is conducted using MATLAB 2024a on a PC with an Intel Core i9-12900K CPU running at 32 GB RAM.

A. Data Description

The performance of the operating mode analysis method is validated using four consecutive years of load data from supervisory control and data acquisition (SCADA) systems in Australia and Singapore. Significant differences in meteorological and geographical features exist between the two countries. The specific data sources are as follows: Australian regional load and related data are from the Australian National Electricity Market [42], in which the samples from 2017 to 2020 are taken; whereas the data of Singapore are obtained from the National Electricity Market of Singapore from 2019 to 2022 [43]. These datasets exhibit high-resolution temporal characteristics and real-world operating conditions, serving as valuable resources for validating the adaptability and generalization capability of the operating mode analysis method across diverse climatic conditions and different power system structures.

B. Operating Mode Analysis

To better reflect the influence of sample size on the identification of power system operating modes, the adaptive threshold AP clustering with DTW is adopted. The intra-class correlation for each cluster of results is calculated according to (8)-(12) to compare the performance between the adaptive threshold AP clustering with DTW and several other clustering methods. The data of Singapore in 2022 and those of Australia in 2020 are used. Table I shows the comparison of clustering results obtained by different methods in two countries, where A represents the AP clustering based on Euclidean distance [44], B represents weighted fuzzy C-means clustering [45], C represents K-means++ with a predetermined number of clusters [24], and D represents the adaptive threshold AP clustering with DTW in this paper.

TABLE I
COMPARISON OF CLUSTERING RESULTS OBTAINED BY DIFFERENT METHODS IN TWO COUNTRIES

Country	Method	$\Gamma_s^{\max}$	$\Gamma_s^{\min}$	Γ_s^{mean}	Γ_s^{median}	Γ_s^{SD}	Number of clusters	Time (s)
Singapore	A	0.9859	0.8178	0.9457	0.9618	0.0380	4	1.66
	B	0.9889	0.7981	0.9446	0.9665	0.0415	4	0.86
	C	0.9734	0.7994	0.9467	0.9559	0.0416	4	2.79
	D	0.9930	0.8057	0.9500	0.9669	0.0347	4	2.99
Australia	A	0.9858	0.7665	0.9237	0.9378	0.0503	5	1.87
	B	0.9897	0.7612	0.9301	0.9427	0.0516	5	0.91
	C	0.9856	0.7582	0.9104	0.9317	0.0581	5	2.83
	D	0.9902	0.7912	0.9311	0.9458	0.0463	5	3.05

As shown in Table I, it is found that in the two countries, method D has the best clustering performance based on the results of five indices in (8)-(12). Although the results confirm that the computational cost of method D is slightly higher than those of methods A and B, its computational time remains within 4 s for the target data scale. This difference is insignificant in practical operating mode analysis. Concurrently, method D eliminates the need for pre-estimating the optimal number of clusters as required in methods B and C, which simplifies the implementation process.

After verifying the performance of the adaptive threshold AP clustering with DTW, it is applied to cluster 30-min-interval load data at five different time scales: six months, one year, two years, three years, and four years in the two countries. The identification results and different characteristics of typical operating modes under different temporal scales are analyzed. Figure 7 shows the identification results of power system operating modes in Singapore and Australia at different time scales. Each point represents a daily operating mode, different colors represent different power system operating modes obtained by clustering, and red pentagrams represent the clustering exemplar of each power system operating mode.

It can be seen from Fig. 7 that when the time scale is longer, the number of power system operating modes obtained through clustering will increase appropriately. Meanwhile, when the sample size is less than 4 years (below 1500 samples), the training time of the adaptive threshold AP clustering with DTW remains less than 7 s, demonstrating excellent real-time analysis capability. Among them, the power system operating modes in Australia are mainly affected by seasonal loads and hydropower generation output, showing obvious seasonality and periodicity in Fig. 7(f)-(h). The daily power system operating modes within seasons are relatively stable, and the changes in power system operating modes exhibit more obvious regularity. As a receiving-end power system, Singapore exhibits severely disordered power system operating mode distribution with highly frequent fluctuations in Fig. 7(a)-(e), leading to the inability to infer future operating modes based on seasonal factors. Using the data-driven method to identify typical operating modes can objectively and reasonably evaluate the power system operating modes from different time scales.

To improve the readability of the results, critical conclusions are quantitatively validated using (8)-(12). Box plots are used to reflect the variability and dispersion of the identification results of the two countries at different time scales. The correlation statistics of clustering results between the two countries at different time scales are presented in Fig. 8.

The analysis reveals two consistent results across both countries despite their different geographical characteristics and data collection periods. ① After excluding outliers, both the median and minimum of the obtained results show similar trends with changes in the time scale, reaching the optimal performance at the time scale of 2 years. ② The distribution of correlation coefficients shows higher sample concentration and lower dispersion at the time scale of 2 years. The clustering results of power system operating modes obtained at this time scale have relatively smaller computational errors compared with other time scales.

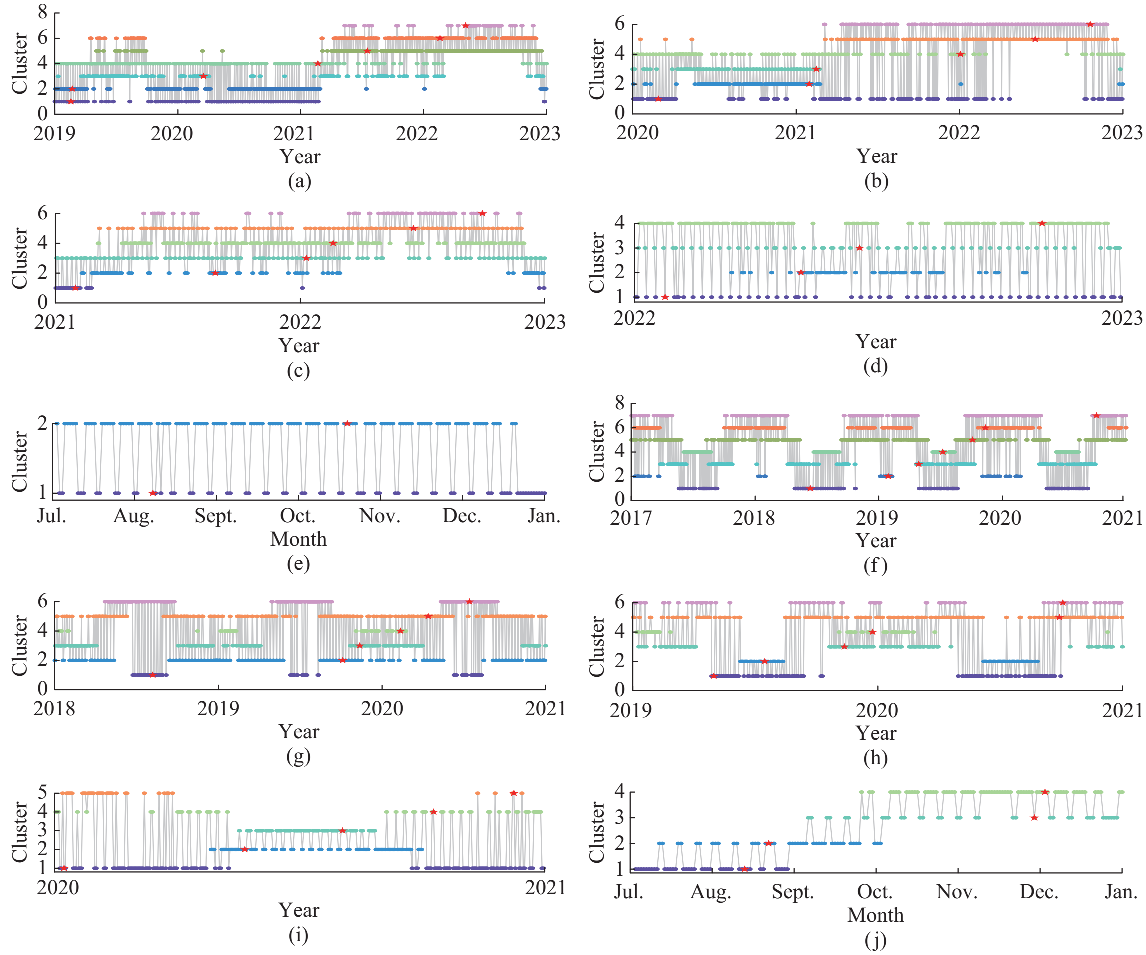

Fig. 7. Identification results of power system operating modes in Singapore and Australia at different time scales. (a) Singapore from 2019 to 2022. (b) Singapore from 2020 to 2022. (c) Singapore from 2021 to 2022. (d) Singapore in 2022. (e) Singapore in second half of 2022. (f) Australia from 2017 to 2020. (g) Australia from 2018 to 2020. (h) Australia from 2019 to 2020. (i) Australia in 2020. (j) Australia in second half of 2020.

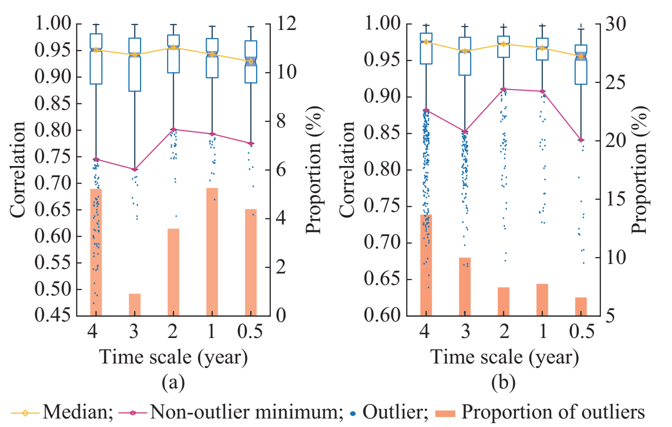

Fig. 8. Correlation statistics of clustering results between two countries at different time scales. (a) Australia. (b) Singapore.

C. Deep Learning for Day-ahead Load Forecasting

To validate the robustness of the PTFN model, partial samples from January, April, July, and October are selected from both countries as test cases. Among these samples, samples before the test date can be used as the training set, with cross validation employed for model training. The ratio of the training set to the validation set is set to be 12:1.

The methodology initiates with feature extraction, where Fig. 9 quantitatively presents the changes in historical Spearman rank correlation coefficients at the corresponding time-points with D-day lead time ($D \in [1, 84]$). As shown in Fig. 9, the historical loads from both countries exhibit distinct periodic fluctuations on a weekly basis, where the input historical loads separated by 7-day intervals demonstrate high correlations prior to the test targets. Due to the differences in electricity consumption behaviors caused by meteorological features, the correlation decreases significantly with the increase of longer time scale in Australia. Combined with the examples shown in Fig. 9(a), the historical loads from 1-day, 7-day, and 14-day intervals preceding the forecasting target are selected as input features for the PTFN model. In contrast, Singapore exhibits relatively stable correlation trends, prompting the selection of 7-day, 14-day, and 21-day historical loads as model inputs in Fig. 9(b).

The PTFN model is compared with the TCN, LSTM, and TCN-LSTM [35] models to validate the reliability of the forecasting results. Supplementary Material A Table SAIII presents the average training accuracy I_{MSE} and total computational cost of each model. Table II presents the load forecasting results based on different models in the two countries. To minimize the randomness as much as possible, all obtained values represent the average of 10 independent training iterations.

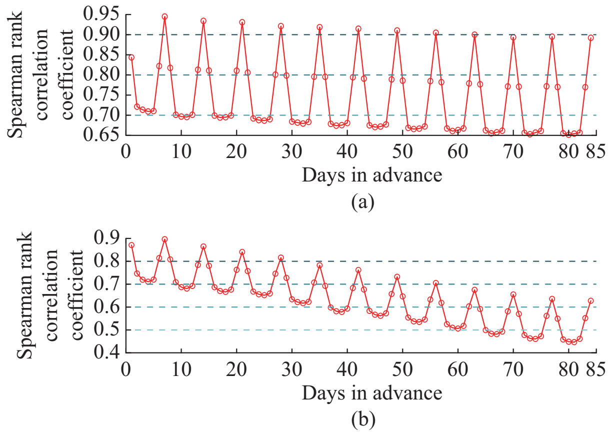

Fig. 9. Changes in historical Spearman rank correlation coefficients. (a) Singapore. (b) Australia.

It can be observed from Table SAIII that the PTFN model achieves significantly higher accuracy than other models with a shorter training time. Additionally, the PTFN model has the lowest forecasting error in 14 of 16 evaluation indices in Table II. Compared with the sub-optimal model in each case, it demonstrates average improvements of 9.37% in I_{RMSE} and 10.28% in I_{MAPE}. Referring to the detailed data in Table II, even in cases without the lowest indices, the average difference between I_{MAPE} of PTFN model and the optimal model remains within 4.31%. The PTFN model achieves the lowest local forecasting errors among all models in the two countries. Referring to Fig. 10, the PTFN model achieves I_{MAPE} predominantly below 3% in the two countries, with both its error distribution center and maximum error being significantly lower than those of other models. Based on the analyses of both global and local forecasting errors, the PTFN model exhibits outstanding forecasting performance across different temporal and spatial scenarios, enabling an accurate day-ahead load forecasting.

TABLE II
LOAD FORECASTING RESULTS BASED ON DIFFERENT MODELS IN TWO COUNTRIES

Country	Date	I_{RMSE} (MW)				I_{MAPE} (%)			
		TCN	LSTM	TCN-LSTM	PTFN	TCN	LSTM	TCN-LSTM	PTFN
Singapore	January 1, 2022-January 7, 2022	159.399	160.052	165.067	154.394	2.128	2.157	2.108	2.091
	April 1, 2022-April 7, 2022	88.641	87.793	86.645	83.245	1.080	1.099	1.055	1.041
	July 1, 2022-July 7, 2022	218.827	246.878	217.845	209.630	2.730	3.195	2.705	2.828
	October 1, 2022-October 7, 2022	115.604	122.776	139.886	106.734	1.561	1.462	1.827	1.382
Australia	January 1, 2020-January 7, 2020	422.984	497.523	447.647	398.796	4.191	4.363	4.118	3.553
	April 1, 2020-April 7, 2020	280.008	249.954	257.004	239.850	2.433	2.359	2.481	2.322
	July 1, 2020-July 7, 2020	393.898	401.788	388.153	349.929	2.954	2.915	2.769	2.892
	October 1, 2020-October 7, 2020	425.337	367.156	396.205	357.806	4.588	3.653	4.179	3.546

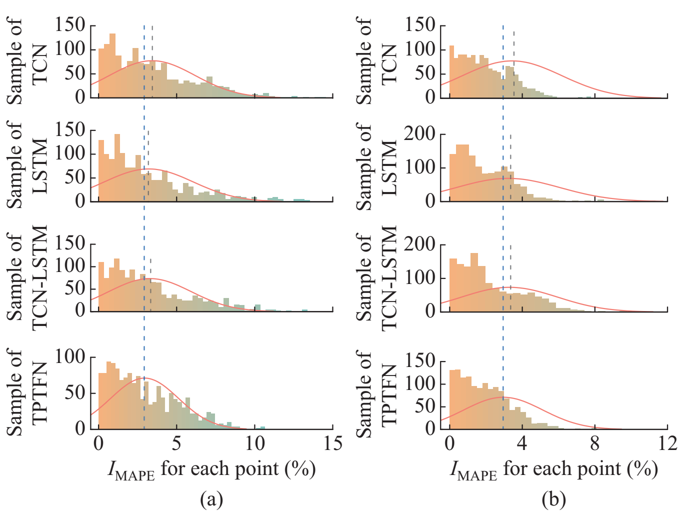

Fig. 10. Comparison of local forecasting errors based on different models in two countries. (a) Australia. (b) Singapore.

D. Power System Operating Mode Identification

1) SHAP-based Feature Extraction

First, the identification results of typical operating modes at a two-year time scale obtained in Section III are used as the baseline. Next, a total of 56 sample sets from different months across two countries forecasted in Section IV are selected as test targets. Finally, SHAP is employed to extract inputs suitable for PTCN-SE in different scenarios from the eight types of features that reflect load level characteristics mentioned in Section V, and Supplementary Material A Fig. SA1 displays the results in different scenarios. As shown in Fig. SA1, following the strategy proposed in [41], this paper selects the features with significantly higher SHAP values as inputs (the features highlighted in red boxes in Fig. SA1). In both countries, four features are selected as model inputs for secondary training, which all include the three-quarter quartile, mean, quartile, and maximum load values.

2) Identification of Future Operating Modes

The training set, validation set, and test set for the PTCN-SE are divided into a ratio of 7:2:1. By comparing the identification accuracies of CNN, TCN, PTCN-SE without feature extraction, and SHAP-PTCN-SE with feature extraction, the rationality of SHAP-PTCN-SE in identifying future operating modes is evaluated. Supplementary Material A Table SA-IV shows the training performances of different models in the two countries. Figure 11 compares the identification results between Singapore and Australia. Table III present the accuracy of different methods to identify future operating modes in two countries.

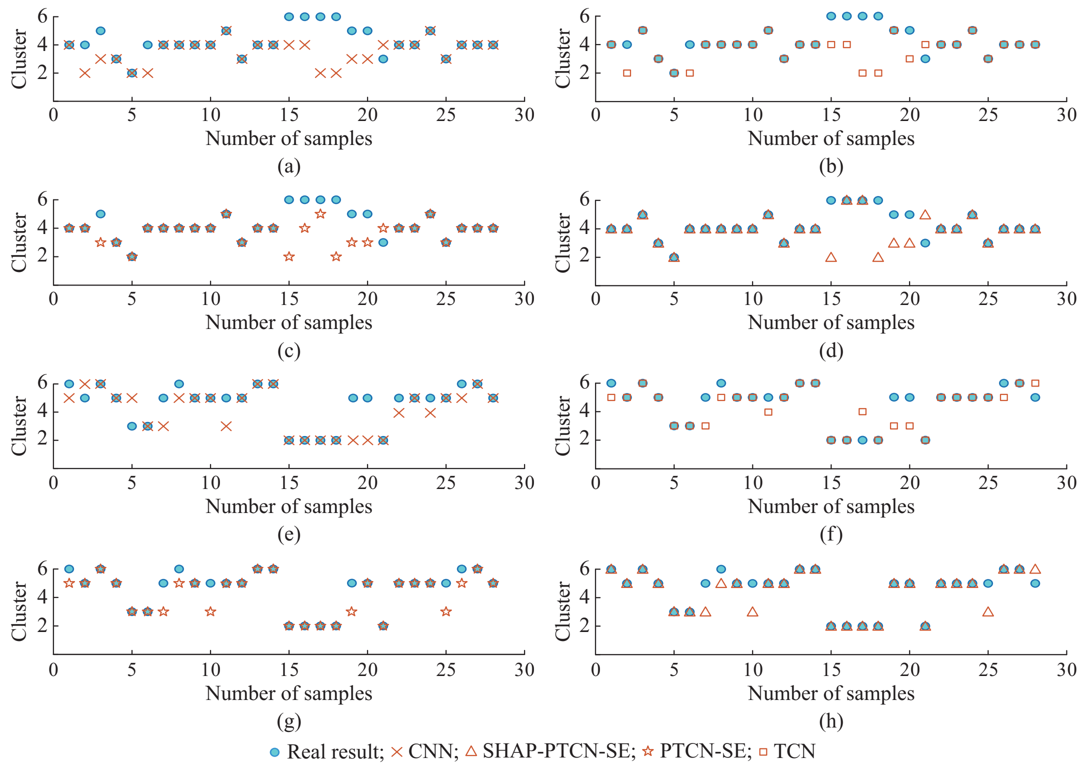

Fig. 11. Comparison of identification results between Singapore and Australia. (a) CNN in Singapore. (b) TCN in Singapore. (c) PTCN-SE in Singapore. (d) SHAP-PTCN-SE in Singapore. (e) CNN in Australia. (f) TCN in Australia. (g) PTCN-SE in Australia. (h) SHAP-PTCN-SE in Australia.

TABLE III
ACCURACY OF DIFFERENT METHODS FOR IDENTIFICATION OF FUTURE OPERATING MODES IN TWO COUNTRIES

Model	Accuracy (%)		Training time (s)
	Singapore	Australia	
CNN	64.29	60.71	4.31
TCN	71.43	67.86	12.19
PTCN-SE	71.43	75.00	8.94
SHAP-PTCN-SE	82.14	82.14	8.16

It can be observed from Table III that SHAP-PTCN-SE achieves significantly higher accuracy than other models. Besides, PTCN-SE can better balance the accuracy and computational efficiency compared with CNN and TCN when identifying future operating modes in two countries. In addition, the introduction of SHAP to extract critical features further enhances the accuracy of PTCN-SE. The results in Table III represent the average accuracy of each model after five training iterations, and the identification accuracy of SHAP-PTCN-SE can still maintain above 82%, which is significantly higher than those of other three models and sufficient to prove its reliability and stability.

In addition, taking the forecasted daily load as input, the process of power system operating mode identification is repeated to determine the mode allocation of 56 samples in two countries. Table IV presents the performance comparison between unsupervised clustering-based and deep learning-based methods. The results from Tables III and IV and Supplementary Material A Table SAIV demonstrate that unsupervised clustering-based method exhibits slightly higher computational efficiency. However, their identification accuracy is at least 20% lower than that of the deep learning-based method. This further demonstrates that deep learning-based method with supervised learning can achieve better results in identifying future operating modes.

TABLE IV
PERFORMANCE COMPARISON BETWEEN UNSUPERVISED CLUSTERING-BASED AND DEEP LEARNING-BASED METHODS

Method	Accuracy (%)		Training time (s)
	Singapore	Australia	
Unsupervised clustering-based	60.71	60.71	7.14
Deep learning-based	82.14	82.14	8.16

3) Impact of Forecasting Errors on Power System Operating Mode Identification

SHAP-PTCN-SE simulates forecasting errors with I_{MAPE} of 1%, 3%, 5%, and 10% by introducing controlled levels of Gaussian noise into the load forecasting results generated by the PTFN model. From the load forecasting datasets with varying error levels, eight types of features reflecting load level characteristics mentioned in Section V are extracted. These features serve as input of SHAP-PTCN-SE. The 56 sample sets from different months across the two countries are still selected as test cases. Through analyzing the accuracy, SHAP-PTCN-SE enables quantitative evaluation of how the accumulation of forecasting errors affects the identification results of future operating modes. The specific results are shown in Fig. 12. As shown in Fig. 12, both countries exhibit a clear declining trend in the accuracy as the forecasting error increases. Error accumulation reveals a significant negative impact on system performance, with accuracy dropping by approximately 30%-40% from the original forecasting error of 0%-10%. As the forecasting error increases from 0% to 5%, the accuracy in the two countries decreases by 17.85% and 21.43%, respectively. This reduction significantly exceeds the decline of accuracy of 14.29% that occurs in both countries when the forecasting error increases from 5% to 10%. These results confirm the marked system sensitivity to forecasting errors. As established above, this further confirms the critical importance of developing more accurate load forecasting methods for identifying future operating modes, as it fundamentally mitigates the impact of error accumulation.

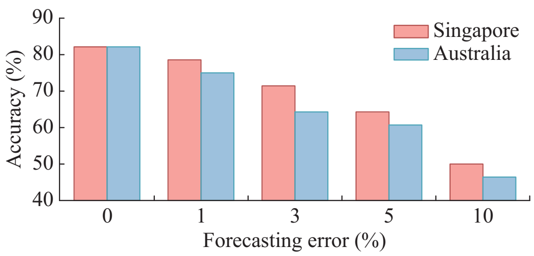

Fig. 12. Specific results of impact of forecasting errors.

VII. Conclusion

To further enhance the accuracy, rationality, and applicability of power system operating mode identification, this paper proposes a data-driven and deep learning-based method integrating operating mode analysis, day-ahead load forecasting, and identification of future operating modes.

1) The adaptive threshold AP clustering with DTW could effectively handle high-dimensional temporal data, demonstrating superior clustering performance compared with traditional methods.

2) The PTFN model incorporating temporal features achieves an excellent balance between accuracy and computational efficiency in day-ahead load forecasting, outperforming several existing models.

3) The SHAP-PTCN-SE for identification of future operating modes features the accumulation of forecasting errors and more robust decisions when addressing the uncertainty of future operating modes, achieved by innovatively integrating feature explanation.

In future work, it is meaningful to incorporate high penetration of renewable energy sources and time-varying operating boundaries while considering load distributions at different nodes.

References

[1] T. Xia, Q. Hou, N. Zhang *et al.*, "Database generation for data-driven power system security assessment under uncertainty," *IEEE Transactions on Power Systems*, vol. 39, no. 5, pp. 6168-6182, Sept. 2024.

[2] Y. Liu, B. Huang, Y. Lin *et al.*, "Optimal SOC headroom of pump storage hydropower for maximizing joint revenue from day-ahead and real-time markets under regional transmission organization dispatch," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 1, pp. 238-250, Jan. 2024.

[3] W. Liang, S. Lin, Y. Yang *et al.*, "Multi-objective robust optimal secure operation model of large-scale power grid with multiple back-to-back voltage source converter based systems considering short-circuit current limitation," *Journal of Modern Power Systems and Clean Energy*, vol. 13, no. 4, pp. 1151-1166, Jul. 2025.

[4] H. Zhang, C. He, H. Shen *et al.*, "Identifying critical transmission cutsets: a mechanism-data-driven method considering static security and transient stability," *International Journal of Electrical Power & Energy Systems*, vol. 172, p. 111346, Nov. 2025.

[5] J. Wang, Y. Jing, C. Zhang *et al.*, "Performance comparison of combined cooling heating and power system in different operation modes," *Applied Energy*, vol. 88, no. 12, pp. 4621-4631, Dec. 2011.

[6] J. Li, J. Zhou, and B. Chen, "Review of wind power scenario generation methods for optimal operation of renewable energy systems," *Applied Energy*, vol. 280, p. 115992, Dec. 2020.

[7] S. Riaz, G. Verbič, and A. Chapman, "Computationally efficient market simulation tool for future grid scenario analysis," *IEEE Transactions on Smart Grid*, vol. 10, no. 2, pp. 1405-1416, Mar. 2019.

[8] R. Ma, M. Zhan, K. Jiang *et al.*, "Transient stability assessment of renewable dominated power systems by the method of hyperplanes," *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp. 3455-3468, Oct. 2023.

[9] H. Zhang, J. Gao, X. Ye *et al.*, "Power system operation modes identification based on data-driven and deep learning," in *Proceedings of the 2023 Panda Forum on Power and Energy*, Chengdu, China, Apr. 2023, pp. 1678-1684.

[10] Y. Wang, N. Zhang, Y. Tan *et al.*, "Combining probabilistic load forecasts," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3664-3674, Jul. 2019.

[11] Y. Wang, Q. Xu, M. Liu *et al.*, "A novel system operation mode with flexible bus type selection method in DC power systems," *International Journal of Electrical Power & Energy Systems*, vol. 103, pp. 1-11, Dec. 2018.

[12] Y. Weng, R. Negi, and C. Faloutsos, "Robust data-driven state estimation for smart grid," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1956-1967, Jul. 2017.

[13] X. Zhang, W. Yan, Z. Lu *et al.*, "Bad data identification for power systems state estimation based on data-driven and interval analysis," *Electric Power Systems Research*, vol. 217, p. 109088, Apr. 2023.

[14] Y. Guo, Y. Li, S. Zhou *et al.*, "A data-driven three-stage adaptive pattern mining approach for multi-energy loads," *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 12, pp. 7455-7467, Dec. 2024.

[15] R. Jain, J. Qin, and R. Rajagopal, "Data-driven planning of distributed energy resources amidst socio-technical complexities," *Nature Energy*, vol. 2, no. 8, p. 17112, Jul. 2017.

[16] N. Zhang, H. Jia, Q. Hou *et al.*, "Data-driven security and stability rule in high renewable penetrated power system operation," *Proceedings of the IEEE*, vol. 111, no. 7, pp. 788-805, Jul. 2023.

[17] J. Zhang, X. Zheng, Z. Wang *et al.*, "Power system sensitivity identification - inherent system properties and data quality," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 2756-2766, Jul. 2017.

[18] J. Cremer, I. Konstantelos, and G. Strbac, "From optimization-based machine learning to interpretable security rules for operation," *IEEE Transactions on Power Systems*, vol. 34, no. 5, pp. 3826-3836, Sept. 2019.

[19] Y. Zhuang, L. Cheng, N. Qi *et al.*, "Problem-driven scenario reduction framework for power system stochastic operation," *IEEE Transactions on Power Systems*, vol. 40, no. 4, pp. 3232-3246, Jul. 2025.

[20] D. Fitiwi, F. Cuadra, L. Olmos *et al.*, "A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty," *Energy*, vol. 90, no. 2, pp. 1360-1376, Oct. 2015.

[21] M. Taheri, M. Abedini, and F. Aminifar, "A novel centralized load shedding approach to assess short-term voltage stability: a model-free using time series forecasting," *IEEE Transactions on Power Delivery*, vol. 38, no. 5, pp. 3076-3083, Oct. 2023.

[22] S. Wogrin, P. Dueñas, A. Delgadillo *et al.*, "A new approach to model load levels in electric power systems with high renewable penetration," *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2210-2218, Sept. 2014.

[23] Y. Li, X. Bai, J. Meng *et al.*, "Multi-level refined power system operation mode analysis: a data-driven approach," *IET Generation, Transmission & Distribution*, vol. 16, no. 13, pp. 2654-2680, Apr. 2022.

[24] Q. Hou, E. Du, N. Zhang *et al.*, "Impact of high renewable penetration on the power system operation mode: a data-driven approach," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 731-741, Jan. 2020.

[25] S. Zhu, H. Ma, L. Chen *et al.*, "Short-term load forecasting of an integrated energy system based on STL-CPLE with multitask learning," *Protection and Control of Modern Power Systems*, vol. 9, no. 6, pp.

71-92, Nov. 2024.
[26] M. Yang, C. Xu, Y. Bai *et al.*, "Investigating black-box model for wind power forecasting using local interpretable model-agnostic explanations algorithm," *CSEE Journal of Power and Energy Systems*, vol. 11, no. 1, pp. 227-242, Jan. 2025.
[27] G. Lu, C. Tsang, H. Yim *et al.*, "Interpretable fault diagnosis for overhead lines with covered conductors: a physics-informed deep learning approach," *Protection and Control of Modern Power Systems*, vol. 10, no. 2, pp. 25-39, Mar. 2025.
[28] B. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, pp. 972-976, Feb. 2008.
[29] H. Li, "Time works well: dynamic time warping based on time weighting for time series data mining," *Information Sciences*, vol. 547, pp. 592-608, Feb. 2021.
[30] T. Teeraratkul, D. O'Neill, and S. Lall, "Shape-based approach to household electric load curve clustering and prediction," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5196-5206, Sept. 2018.
[31] C. Yu, G. Yan, C. Yu *et al.*, "A multi-factor driven spatiotemporal wind power prediction model based on ensemble deep graph attention reinforcement learning networks," *Energy*, vol. 263, p. 126034, Jan. 2023.
[32] K. Chen, K. Chen, Q. Wang *et al.*, "Short-term load forecasting with deep residual networks," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3943-3952, Jul. 2019.
[33] S. Li, Y. AbuFarha, Y. Liu *et al.*, "MS-TCN++: multi-stage temporal convolutional network for action segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 6, pp. 6647-6658, Jun. 2023.
[34] D. Korkmaz, "SolarNet: a hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting," *Applied Energy*, vol. 300, p. 117410, Oct. 2021.
[35] F. Huo, Y. Chen, W. Ren *et al.*, "Prediction of reservoir key parameters in 'sweet spot' on the basis of particle swarm optimization to TCN-LSTM network," *Journal of Petroleum Science and Engineering*, vol. 214, p. 110544, Jul. 2022.
[36] D. Hendrycks and K. Gimpel. (2016, Jun.). Gaussian error linear units (GELUs). [Online]. Available: https://arxiv.org/abs/1606.08415
[37] N. Yang, J. Hao, Z. Li *et al.*, "Data-driven decision-making for SCUC: an improved deep learning approach based on sample coding and seq2seq technique," *Protection and Control of Modern Power Systems*, vol. 10, no. 2, pp. 13-24, Mar. 2025.
[38] T. Zhang, T. Zhu, K. Gao *et al.*, "Balancing learning model privacy, fairness, and accuracy with early stopping criteria," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 9, pp. 5557-5569, Sept. 2023.
[39] J. Hu, L. Shen, S. Albanie *et al.*, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2011-2023, Aug. 2020.
[40] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427-437, Jul. 2009.
[41] J. Fu, Y. Sun, Y. Li *et al.*, "An investigation of photovoltaic power forecasting in buildings considering shadow effects: modeling approach and SHAP analysis," *Renewable Energy*, vol. 245, p. 122821, Jun. 2025.
[42] L. Bardwell, L. Blackhall, and M. Shaw, "Emissions and prices are anticorrelated in Australia's electricity grid, undermining the potential of energy storage to support decarbonisation," *Energy Policy*, vol. 173, p. 113409, Feb. 2023.
[43] T. Loi and G. Jindal, "Electricity market deregulation in Singapore – initial assessment of wholesale prices," *Energy Policy*, vol. 127, pp. 1-10, Apr. 2019.
[44] P. Li, W. Gu, L. Wang *et al.*, "Dynamic equivalent modeling of two-staged photovoltaic power station clusters based on dynamic affinity propagation clustering algorithm," *International Journal of Electrical Power & Energy Systems*, vol. 95, pp. 463-475, Feb. 2018.
[45] Z. Wu, B. Wang, and C. Li, "A new robust fuzzy clustering framework considering different data weights in different clusters," *Expert Systems with Applications*, vol. 206, p. 117728, Nov. 2022.

Chuan He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Sichuan University, Chengdu, China, in 2011, 2014, and 2018, respectively. He is currently an Associate Professor with the College of Electrical Engineering, Sichuan University. His research interest includes robust optimization on power system operation with renewable energy.

Honghao Zhang received the B.S. degree in automation from Northeast Electric Power University, Jilin, China, in 2019, and the M.S. degree in electronics and communication engineering from Ningxia University, Yinchuan, China, in 2021. He is currently working toward the Ph.D. degree in electrical engineering from Sichuan University, Chengdu, China. His research interest includes application of data-driven method in complex power system operation.

Lu Nan received the B.S. and Ph.D. degrees in electrical engineering from Sichuan University, Chengdu, China, in 2016 and 2020, respectively. She is currently an Associate Professor with the College of Electrical Engineering, Sichuan University. Her research interest includes power system operation.

Jiawei Hu received the B.S. degree in electrical engineering from Sichuan University, Chengdu, China, in 2024, where he is currently pursuing the M.S. degree. His research interests include meta-reinforcement learning method for power system security and stability analysis.

Tianqi Liu received the B.S. and M.S. degrees in electrical engineering from Sichuan University, Chengdu, China, in 1982 and 1986, respectively, and the Ph.D. degree in electrical engineering from Chongqing University, Chongqing, China, in 1996. She is currently a Professor with the College of Electrical Engineering, Sichuan University. Her main research interests include power system analysis and stability control.

Biao Wang received the B.S. degree in electrical engineering from Xi'an University of Technology, Xi'an, China, in 2007, and the M.S. degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2010. He is currently an Engineer in State Grid Sichuan Electric Power Company, Chengdu, China. His current research interests include power system planning and operation.

Jian Gao received the B.S. degree in electrical engineering from Chongqing University, Chongqing, China, in 1998, and the M.S. degree in software engineering from University of Electronic Science and Technology of China, Chengdu, China, in 2015. He is currently the Deputy Director of Sichuan Power Dispatching and Control Center, State Grid Sichuan Electric Power Company, Chengdu, China. His research interests include power grid dispatching and operation.

Xi Ye received the B.S. degree from Sichuan University, Chengdu, China, in 2010, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015. She is currently an Engineer in State Grid Sichuan Electric Power Company, Chengdu, China. Her research interest includes power system planning.