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

RISK DETECTION OF OPERATIONAL FAULTS IN DISTRIBUTION GRID EQUIPMENT BASED ON SUPPORT VECTOR MACHINES AND RISK CHARACTERIZATION

Ning Gao,* Chao Wu,* Ling Liu,* Fei Yang,* Na Liu,* and Chen Shen*

Abstract

The distribution network is a complex dynamic system comprising various types of equipment and involving extensive operational data across different time periods, which poses challenges for accurate real-time detection of operational failure risks in such equipment. To obtain full lifecycle operational data of dynamic distribution network equipment and accurately detect its failure risk in real time, a method based on a support vector machine (SVM) and risk characteristics is proposed for detecting operational failure risks in distribution network equipment. Digital twin technology is employed to construct a dynamic model of the distribution network. The updated dynamic model, combined with load analysis and equipment condition awareness, enables accurate real-time acquisition of lifecycle operational data from distribution network equipment, thereby providing comprehensive and accurate data support for subsequent detection. The Randomized Lasso algorithm is used to extract key risk characteristics associated with equipment operational failures from the data, forming a key risk feature set. Using this feature set as input, a least squares support vector machine (LSSVM) model is constructed to detect and provide early warnings for operational failure risk categories of distribution network equipment. Experimental results demonstrate that the proposed method can effectively detect all categories of operational failure risks in the test distribution network equipment. The detection results are consistent with actual conditions and allow early risk warning based on the detection outcomes. The Kappa coefficient of the detection results reaches above 0.9, the coverage rate is close to 100%, and the average bandwidth remains below 0.8, indicating high detection accuracy and coverage.

* State Grid Hebei Province Electric Power Co., Ltd. Zhengding County Power Supply Branch, Shijiazhuang, Hebei, 050800, China; e-mail: cifangao0e0@163.com, chaoluchuovv0@163.com, tuilinanx4@163.com, doufei42590839508@163.com, naohei004667@163.com, xielashedvz@163.com
Corresponding author: Ning Gao

Recommended by Hasmat Malik
(DOI:10.2316/J.2025.203-0615)

Key Words

Support vector machine; Risk characteristics; Distribution network equipment; Operation failure risk; Digital twins; Randomized Lasso algorithm

1. Introduction

Distribution network equipment operation faults of various types included disconnection, short circuit, grounding, overload, and other line faults, as well as faults in transformers, circuit breakers, distribution cabinets, and other equipment [1]. When these faults occurred during the operation of the distribution network, power outages may result, causing inconvenience and economic losses to users [2]. Simultaneously, distribution network equipment may undergo aging and damage, leading to reduced performance and service life, which adversely affects the safe and stable operation of the entire power grid [3]. Fault risk detection for distribution network equipment was essential to avoid these problems and ensure power grid stability. To identify and address fault risks during distribution network operation, appropriate methods were required to guarantee detection accuracy and efficiency [4]. Through real-time monitoring and detection of distribution network equipment, potential faults could be identified and resolved promptly to prevent their occurrence or escalation, thereby ensuring grid safety and stability [5]. Based on real-time monitoring results, equipment failure points could be quickly located during fault events, emergency response procedures could be activated, and effective measures could be taken to isolate fault areas, minimizing the impact on grid-wide operation. This process provided critical data support for future grid maintenance projects requiring precision investment [6]. Additionally, timely fault detection and repair helped to shorten outage durations, reduce affected areas, improve power supply reliability, and decrease economic losses experienced by users due to power outages [7].

At present, research on this aspect has included the series DC (Direct Current) arc fault detection method [8] proposed by Xiu, Y, and others. This method used the graphical information of the target arc to build a microgrid linear system model. With the power supply and load of the model as inputs, an unknown input observer network was constructed, through which fault lines could be identified. This method employed an unknown input observer to detect series DC arc faults in DC microgrids, accurately capturing weak signal changes when an arc fault occurs, thereby improving detection sensitivity and accuracy. However, the detection performance of this method depended heavily on the accuracy of the system model. If a significant deviation existed between the system model and the actual system, the detection accuracy and reliability may be compromised. Current research in this field has also included the distribution network fault location method [9] proposed by Zhao, M. et al., extracting measured feature vectors from distribution lines and using these vectors as inputs to a deep convolutional neural network model for classifying distribution network line faults. This method enabled automated processing and reduced the need for manual intervention and human errors. However, in practical applications, this method often suffers from insufficient training data, and the deep convolutional neural network model used is overly complex, making it prone to overfitting and exhibiting poor generalization capability. Noushin, D. et al. studied an early fault diagnosis method for distribution lines [10]. They used CoroCam 6D2 ultraviolet imaging technology to acquire ultraviolet and visible light video data of distribution lines. The observation distance, gain of the imager, air pressure, and humidity were considered as effective parameters for characterizing the discharge area of distribution lines in the video data. Video frames were extracted at a specified rate based on the nominal voltage of the line. The extracted video frames were input into a Faster R-CNN (Faster Region-based Convolutional Neural Network) model. Line faults were detected in each video frame, enabling fault diagnosis. The UV (Ultraviolet) imaging technology used in this method did not require direct contact with operating equipment, did not affect the normal operation of the power system, and eliminated the need for power outage detection, thereby improving convenience and security. However, environmental factors such as temperature, humidity, and altitude influenced the corona discharge intensity in UV imaging technology. Absolute calibration of the detection results was difficult in practical applications, which could lead to significant variations under different environmental conditions, affecting diagnostic accuracy. Ahmadi, H. et al. studied a classification method for transformer short-circuit faults [11]. This method classified transformer short-circuit faults by constructing a three-phase voltage-current difference trajectory curve of the transformer. The trajectory curve changes and the number of terminals were monitored in real-time, combined with the finite element method. The finite element method accurately solved the electromagnetic field distribution and electromagnetic parameters inside the transformer through fine mesh generation and numerical calculation, aiding in the precise diagnosis of fault types and locations. However, the finite element method required fine mesh generation in the solution space, necessitating the solution of a large number of algebraic equations. This resulted in substantial computational burden, high resource requirements, and increased time costs, thereby reducing the timeliness of transformer fault classification and detection.

Digital twin technology was an innovative method based on digital technology. It created a corresponding virtual representation, called a "digital twin", by implementing comprehensive digital modeling of physical entities or systems [12]. A digital twin refers to the integration of multi-disciplinary, multi-physical quantity, multi-scale, and multi-probability simulation processes. It utilized physical models, sensor updates, operational history, and other data to complete mapping in virtual space, thereby reflecting the entire life cycle process of the corresponding physical equipment [13]. SVM was a powerful supervised learning algorithm mainly used for classification and regression tasks [14]. As a generalized linear classifier, it performed binary classification of data according to supervised learning, with its decision boundary being the maximum-margin hyperplane for learning samples [15]. SVM implemented classification by finding the optimal boundary between data points. This boundary, called a hyperplane, maximized the margin between different categories. SVM mapped data to a high-dimensional space through kernel functions, enabling it to handle nonlinear classification problems. SVM offered advantages including high efficiency, capability to handle nonlinear problems, strong robustness, and good interpretability [16].

In summary, this study combines digital twin technology and SVM to investigate a risk detection method for distribution network equipment operation faults based on risk characteristics.

2. Distribution Network Equipment Operation Fault Risk Detection Methods

2.1 Dynamic Modeling of Distribution Networks

The digital twin technology is used to construct a "time-space-state" multidimensional distribution network dynamic model. This model integrates temporal, regional, meteorological, future grid planning information, and equipment operational data, covering "planning-construction-operation" multi-temporal aspects [17]. The model realizes dynamic simulation of the distribution network's entire lifecycle and provides more comprehensive and accurate data support for detecting operational fault risks in distribution network equipment. The dynamic model of the distribution network constructed based on digital twin technology is shown in Figure 1.

The dynamic model of the distribution network construction mainly comprises three parts: the physical model, the Bus/Breaker model, and the calculation model,

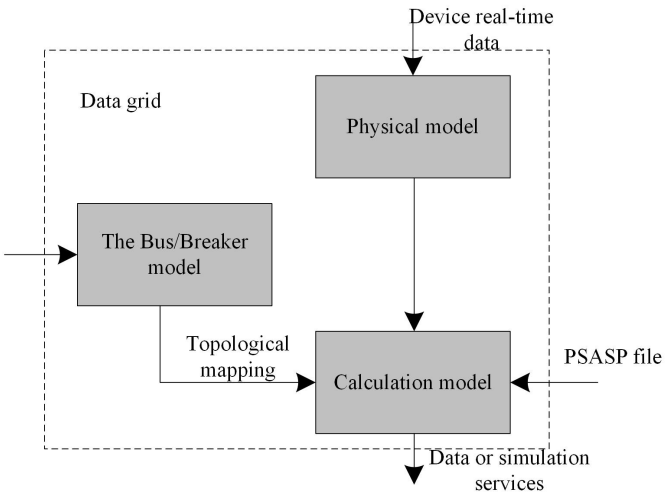


Figure 1. Dynamic Model of Distribution Network

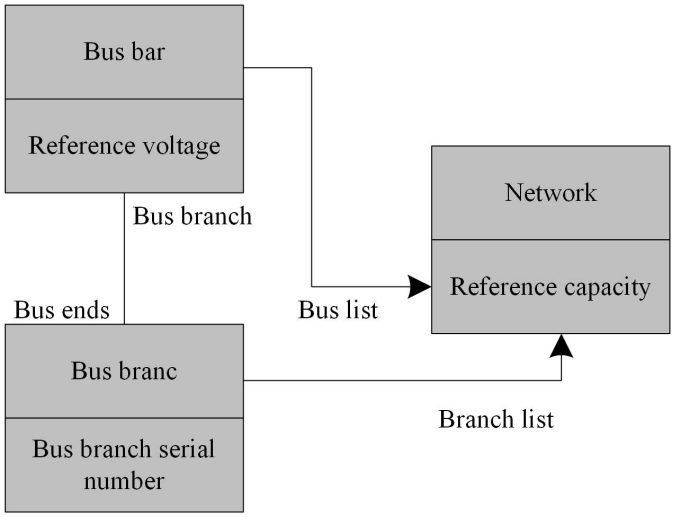


Figure 2. Calculation Model of Distribution Network Dynamic Model

along with the topological mapping relationships among these components [18]. The description of each part is as follows:

(1) *Physical model:* The distribution network node/switch model describes the parameters and connection relationships of physical equipment in the distribution network, including switches, transformers, and other devices. This model corresponds to the measurement information from the data acquisition and monitoring system, as well as the switch setup information. It maps the real-time operational status of the actual physical equipment in the distribution network.

(2) *Bus/Breaker model:* The online analysis data model contains state estimation results. This model is generated by the control cloud system from D5000 or QS files in CIM/E format.

(3) *Calculation model:* The distribution network bus/branch model is generated from the PSASP file, which contains the basic data of distribution network equipment, such as buses, branches, transformers, and generators. This model provides real-time distribution network power flow section data and enables simulation analysis of distribution network equipment based on in-memory computing. The calculation model is shown in Figure 2.
The calculation model employs an object-oriented modeling approach. The distribution network is represented as a graph object. Bus objects serve as vertex objects that connect groups of branch objects. Branch objects function as edge objects, with each branch connecting two bus objects at its terminals. The network acts as the object container for the distribution network diagram, allowing the definition of bus vertex objects and branch edge objects. These components collectively form a network diagram model suitable for distribution network simulation calculations and analysis.

(4) *Mapping relationship:* The topological analysis establishes mapping and automatic collaborative updating relationships among the physical model, Bus/Breaker model, and computing model. To obtain an accurate dynamic distribution network model, automatic updates to the calculation model are required when the operational status of the distribution network's physical equipment changes. The calculation model update implementation scheme uses physical model updates as triggering events. The complex event-processing engine subscribes to these model update events and coordinates corresponding calculation model updates. Physical model updates incorporate load analysis and state perception of distribution network equipment. These updates fully consider temporal, regional, climatic, and other external factors to implement dynamic adjustments. This process produces a dynamically updated distribution network model that maintains consistency with the target distribution network throughout its lifecycle and accurately mirrors the real-time operational status of the distribution network and equipment. The physical model updates in real-time when reading changed operational measurement information from distribution network equipment. Concurrently, the calculation model updates through topology mapping, achieving real-time distribution network "mirroring". This enables accurate real-time acquisition of equipment parameters, operational statuses, defects, potential hazards, weather data, historical fault records, and maintenance records. These capabilities establish foundations for subsequent distribution network equipment operational fault risk detection.

During the process of updating the physical model, external influencing factor data and equipment state perception data are relatively easy to obtain through measuring devices and sensor equipment. Load analysis data of distribution network equipment proves more difficult to acquire. This study combines the exponential smoothing method to obtain load prediction data for distribution network equipment. The exponential smoothing method extrapolates from historical load data of distribution network equip-

ment. Equipment load forecast data are obtained through this method. The exponential smoothing method essentially functions as a moving average technique. It primarily depends on the algorithm assigning weights to historical load data. The weights for recent historical load data of distribution network equipment are set greater than those for long-term historical data. The quadratic exponential smoothing algorithm is selected to implement probabilistic load forecasting for distribution network equipment. The load forecasting process using the quadratic exponential smoothing algorithm follows these steps:

(1) For raw historical load data U of distribution network equipment, implement primary exponential smoothing to obtain a primary exponential smoothing result $U_t^{(1)}$ for such historical load data.

(2) On this basis, after the quadratic exponential smoothing operation, the quadratic exponential smoothed series $U_t^{(2)}$ of such historical load data is obtained. The equation for the quadratic exponential smoothing operation is:

$$\begin{cases} U_t^{(1)} = U_{t-1}^{(1)}(1-\phi) + \phi a_t \\ U_t^{(2)} = U_{t-1}^{(2)}(1-\phi) + \phi U_t^{(1)} \end{cases} \tag{1}$$

In Equation (1), ϕ denotes the smoothing coefficient, where $0 < \phi < 1$, and a_t indicates the load value of the distribution grid equipment at time period t, with $t = 1, 2, \ldots, T$, where T indicates the total number of periods for which equipment loads are forecasted. U_t represents the load data for the t-th period, which is used to predict the value of the equipment load for the $(\ t+1\ )$-th period.

(3) Calculate the intercept and slope of the load forecasting line in the quadratic exponential smoothing operation, both of which can be expressed as:

$$\begin{cases} c_t = 2U_t^{(1)} - U_t^{(2)} \\ d_t = \phi\left(U_t^{(1)} - U_t^{(2)}\right)/(1-\phi) \end{cases} \tag{2}$$

In Equation (2), c_t and d_t represent the intercept and slope of the load forecasting line.

(4) Based on the calculation results obtained from Equation (2), the load forecast value of the distribution network equipment for the next K years can be obtained.

$$A_{T+K} = c_T + d_T K, K = 1, 2, ..., M \tag{3}$$

In Equation (3), M represents the projected future lifespan of the distribution network equipment.

2.2 Extraction of Key Risk Characteristics of Distribution Network Equipment Operation Failure Based on Randomized Lasso Regression.

Through the analysis and preprocessing of operational data obtained from the dynamic model of the distribution network, the fault characteristics of distribution network equipment can be classified into four categories: fault factors, external factors, self-factors, and operational factors, comprising a total of 24 characteristics as shown in Table 1.

Combined with the Randomized Lasso regression algorithm, the key features of the operational failure risk of distribution network equipment are extracted from these operational failure features to form a key risk feature set. Lasso regression is an extension of linear regression. It achieves the selection of key risk characteristics of equipment operational failure by adding an L1 regularization term, which can reduce the coefficients of some risk characteristics to zero while fitting data to accomplish the purpose of selecting key risk characteristics [19]. The operational fault characteristics of distribution network equipment belong to high-dimensional features, and the computational complexity of Lasso regression in high-dimensional feature operations may be high. It is necessary to introduce randomization to reduce the dimensionality of such high-dimensional characteristics and improve the efficiency of selecting the key risk characteristics of the equipment. For this reason, this study selects the Randomized Lasso regression algorithm, which projects the high-dimensional features of distribution network equipment operational faults into a low-dimensional space by constructing a random matrix to generate a random projection feature matrix. In addition, the Lasso regression model is applied to select key risk characteristics of equipment in the low-dimensional space after random projection [20]. The key risk characteristics extraction process of distribution network equipment operational failure based on the Randomized Lasso regression algorithm is as follows:

(1) Let the multiple linear regression model satisfying the classical assumptions be as follows:

$$B = \mu + \alpha s \tag{4}$$

In Equation (4), α denotes the objective function of the regression model; μ represents the coefficients of each residual in the model, which are independent and identically normally distributed;s denotes the fault characteristics of the distribution network equipment operation. Thus, the response variable of this regression model is:

$$B \sim L\left(\alpha s, \mu^2 Q\right) \tag{5}$$

In Equation (5), L indicates the number of response variables; Q denotes the unit matrix. The objective function of this underlying regression model is:

$$\min_{\alpha}\left(\frac{|B-\alpha s\|_2^2}{2}\right) \tag{6}$$

By taking the objective function of Equation (6) to its extremes, the ordinary least squares estimation is obtained as follows:

$$\alpha' = s's^{-1}s'B \tag{7}$$

(2) Assuming that the rows and columns in the distribution network equipment operation fault risk characterization matrix are denoted as w and o, respectively, if

Table 1
Classification of Distribution Network Equipment Operation Fault Characteristic Variables

Feature type	Feature name	Characteristic variable
Operational factor	Monthly average feeder load	s1
	Average time of fuse operation	s2
	Average operation time of sectional cable	s3
	Average operating time of load switch	s4
	Average operation time of branch lines	s5
	Average operation time of transformer	s6
	Average service time of insulated conductors	s7
	Number of transformers	s8
	Number of fuses	s9
Self factor type	Length of segmented insulated conductors	s10
	Number of segmented insulated wires	s11
	Cable length	s12
	Number of feeder branch lines	s13
External factor type	Mean monthly temperature	s14
	Mean monthly precipitation	s15
	Extreme high temperature	s16
	Extreme low temperature	s17
	Month gale day grading	s18
	Monthly thunderstorm day classification	s19
	Daily precipitation classification	s20
Fault factor type	Total loss of load	s21
	Monthly number of failures	s22
	Feeder substation	s23
	Down time	s24

$w < o$, it indicates that the sample size of distribution network equipment operation fault risk characteristics is smaller than the number of explanatory variables. At this time, rand $(s's) \leq \text{rank}(s) \leq w$, where $(s's)$ is a squares of o line o columns. Due to the non-existence of ($s's$) in this case, ordinary least squares estimation cannot be implemented. Therefore, by adding the L1 regularization term to Equation (6), the objective function of the Lasso regression model is constructed as follows:

$$\min_{\alpha}\left(\frac{\|B-\alpha s\|_2^2}{2}+\frac{\gamma\|\alpha\|_1}{2}\right) \tag{8}$$

Here, γ denotes the penalty parameter, which is positively related to the penalty intensity. $\|\alpha\|_1$ denotes the L1-norm, satisfying $\|\alpha\|_1 = |\alpha_1| + |\alpha_2| + \cdots + |\alpha_z|$. In Equation (8), the first term within the parentheses is the loss function, which measures the regression model's fitting performance on the equipment operation data; the second term is the penalty function, which compresses the coefficients of some insignificant equipment risk features to zero.

(3) Therefore, the key risk feature selection in the Lasso regression model can be determined by whether the risk feature coefficient is zero.

$$\hat{S} = \{i : \alpha_i' \neq 0\} \tag{9}$$

In Equation (9), $\hat{S} \subset \{1, \cdots, q\}$ represents a random projection feature set that enters the Lasso regression model after feature selection, where q denotes the number of features in the set.

In the random projection feature matrix of Equation (9), by optimizing Equation (8), the Randomized Lasso regression constraint equivalent to Equation (8) can be obtained as follows:

$$\begin{cases} \min_{\alpha}\left(\frac{\|B-\alpha s\|_2^2}{2}\right) \\ \text{s.t } \|\alpha\|_1 \leq \eta \end{cases} \tag{10}$$

In Equation (10), η denotes the constraint parameter corresponding to the penalty parameter γ, which is negatively correlated with the penalty intensity. The optimal solution, i.e., the key risk characteristics of the operational faults of the distribution network equipment, can be obtained by utilizing Equations (8) and (10), and the key risk characteristic set S is constructed. The key risk features are ranked by assigning weights to each key risk feature in the resulting set.

2.3 Distribution Network Equipment Operation Fault Risk Detection Based on Key Risk Characteristics and LSSVM

Taking the set of critical risk features for operational failures of distribution network equipment after the weight ordering in the previous section, denoted as S, as an input, the LSSVM model is used to construct a detection and early warning model for the operational failure risk of

distribution network equipment to achieve real-time and accurate monitoring and warning of such risks. In the LSSVM model, the least-squares linear system is employed as the loss function to accelerate the solution process of nonlinear function estimation and improve the operational efficiency of the overall risk detection and early warning model. The key risk feature set S of distribution network equipment is divided into N groups of feature samples, i.e., $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$. Through LSSVM, the regression problem of SVM for these feature samples is transformed into a convex optimization problem as described in the following equation:

$$\begin{cases} \min\limits_{\vartheta, g, f} \delta(\vartheta, f) = \dfrac{\|\vartheta\|^2}{2} + \dfrac{\lambda \sum_{j=1}^{N} f_j^2}{2} \\ \text{s.t. } \vartheta^{\mathrm{T}} \sigma(x_j) + g = y_j - f_j, j = 1, 2, \cdots, N \end{cases} \tag{11}$$

In Equation (11), f_j denotes the error variable; λ denotes the regularization parameter, which adjusts the accuracy and generalizability of the LSSVM model to avoid overfitting; ϑ denotes the normal vector to the hyperplane; g indicates the bias term; δ represents the hyperplane; and σ denotes the nonlinear transformation function of the LSSVM model. To avoid the computational challenges in high-dimensional space, the Lagrangian method is generally employed to solve the problem, and the solution equation is as follows:

$$\text{Lagrange}(\vartheta, g, f, \beta) = \delta(\vartheta, f) - \sum_{j=1}^{N} \beta_j \left\{ \vartheta^{\mathrm{T}} \sigma(x_j) + g + f_j - y_j \right\} \tag{12}$$

In Equation (12), $\beta = [\beta_1, \beta_2, \cdots, \beta_N]^{\mathrm{T}}$ denotes the Lagrange multiplier. Combined with the duality theory, Equation (12) can be transformed into the following matrix form:

$$\begin{bmatrix} Q & 0 & 0 & -\sigma \\ 0 & 0 & 0 & 1^{\mathrm{T}} \\ 0 & 0 & \lambda Q & -Q \\ \sigma & 1 & Q & 0 \end{bmatrix} \begin{bmatrix} \vartheta \\ g \\ f \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix} \tag{13}$$

After eliminating the variables ϑ and f in the equation, the following system of equations can be obtained:

$$\begin{bmatrix} 0 & 1^{\mathrm{T}} \\ 1 & \Omega + \lambda^{-1} Q \end{bmatrix} \begin{bmatrix} g \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{14}$$

In Equation (14), $\Omega_{i,j} = \sigma(x_j)^{\mathrm{T}} + \sigma(x_i), i = 1, 2, \cdots, N$, and $y = (y_1, y_2, \cdots, y_N)^{\mathrm{T}}$.

After calculating the values of parameters g and β using Equation (14), the estimated prediction function for the feature sample x is:

$$h(x) = \sum_{j=1}^{N} g + P(x, x_j) \beta_j \tag{15}$$

In Equation (15), $P(x, x_j)$ denotes the kernel function. The kernel function used in this paper is the Gaussian radial basis function, expressed as:

$$P(x, x_j) = \exp\left[\frac{-\|x - x_j\|^2}{2v^2} \right] \tag{16}$$

In Equation (16), v denotes the kernel width of the kernel function.

The fault risk detection process of the distribution network equipment operation based on the LSSVM model is as follows:

(1) The LSSVM model is constructed using the least squares method;
(2) The key risk feature set S of distribution network equipment extracted by the Randomized Lasso regression algorithm is divided into N groups of feature samples $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, which are used as input to the LSSVM model to train the model;
(3) During the training process, using the selected Gaussian radial basis kernel function $P(x, x_j)$ with the regularization parameter λ, optimize the performance of the LSSVM model;
(4) The real-time operation data of the distribution network equipment collected by the distribution network dynamic model are input into the trained and optimized LSSVM model, and the operational failure risk classification results of the equipment are obtained;
(5) Based on the results of risk classification, the real-time operating status of distribution network equipment is determined, potential operational failure risks are timely detected, and early warnings are issued to realize equipment operational failure risk detection and warning.

Multi-parameter estimation significantly improves the lag problem of traditional single-parameter control by integrating multidimensional real-time data. Traditional methods only monitor a single parameter, while multi-parameter estimation simultaneously analyzes various characteristics such as equipment load, environmental factors, and historical failures, forming a dynamic collaborative evaluation system. The multidimensional model established through digital twin technology achieves real-time updates of topology mapping relationships, enabling the system to perceive the coupling risk of temperature spikes and insulation aging in advance. The random regression method further screens key features, combined with the kernel function of SVM, to reduce the response delay from minutes to seconds. This multi-dimensional parameter coupling analysis effectively overcomes the perception blind spot of single-parameter systems for composite faults. The system achieves early warning and precise positioning of potential risks by comprehensively considering the mutual influence of various factors, significantly improving the reliability and safety of the power system.

3. Analysis of Experimental Results

Using the distribution network in Zhengding County as an example, this method is employed to detect the operational failure risk of distribution network equipment. The

application effectiveness of this method is evaluated based on actual detection results. Zhengding County belongs to Shijiazhuang City, Hebei Province, and is located in the central area of Shijiazhuang, with a total area of 486 km^2. It features a temperate monsoon climate with distinct continental monsoon characteristics and four distinct seasons. The distribution network in Zhengding County is operated by the State Grid Zhengding County Power Supply Company and serves as part of a comprehensive demonstration project for a new type of power system that aims to promote the transformation and upgrading of traditional energy resources through an innovation-driven development strategy. It enhances the capabilities of "station, line, transformer, household" hierarchical classification and integrated "source, network, load, storage" scheduling, enabling the visualization, measurement, adjustment, and control of second-level photovoltaic distributed power supply, energy storage, and user load. The Zhengding County distribution network is shown in Figure 3.

Figure 3. Experimental Positive Definite County Distribution Network

A set of experimental visualization platforms is constructed to clearly demonstrate the application effectiveness of this method. The platform is based on a dynamic map of the distribution network in Zhengding County, and the multi-dimensional and multi-state operational data of experimental distribution network equipment are superimposed in a hierarchical manner to realize the visualization of the results of detecting and early warning of the operational failure risk of experimental distribution network equipment using the method proposed in this paper in real time. The effectiveness of the constructed experimental distribution network equipment detection visualization platform is shown in Figure 4.

The experimental dataset consists of collected and organized data related to the equipment operation of the experimental distribution network from January 2021 to December 2022. The experimental dataset is randomly divided into training and test datasets, where the training dataset is used to train and continuously optimize the risk detection model of the proposed method, while the test dataset is employed to evaluate the risk detection and early warning effectiveness of the method presented in this study. First, the method used in this study utilizes digital twin technology to construct a dynamic model of the experimental distribution network, and the construction result is shown in Figure 5.

Figure 4. Experimental Visualization Platform

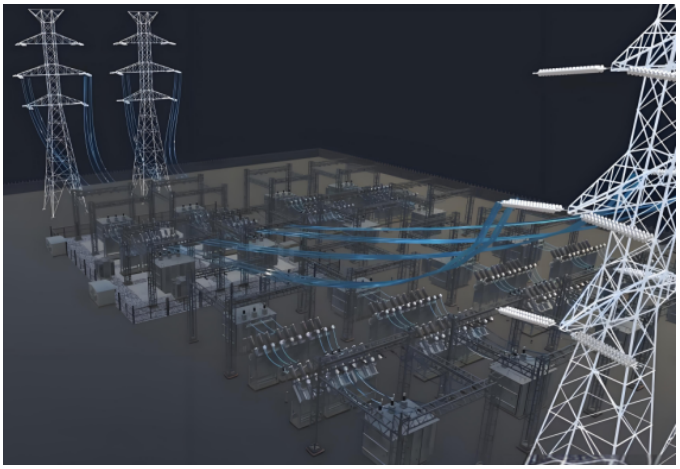


Figure 5. Digital Twin-based Dynamic Model of the Experimental Distribution Network Constructed by the Proposed Method.

The key risk characteristics of the distribution network equipment operational failures extracted using the methodology of this study are listed in Table 2.

It can be seen from Table 2 that the number of key risk characteristics of operational failure of distribution network equipment extracted by the method in this study is 19, and five risk characteristics of s10, s11, s12, s13, and s 15 corresponding to the length of sectional insulated conductors, the number of sectional insulated conductors, cable length, the number of feeder branch lines, and the monthly average precipitation are excluded. The weight values of the 19 extracted key risk characteristics are shown in Figure 6.

Table 2
Key Risk Characteristics of Distribution Network Equipment Operation Failure Extracted by the Proposed Method

Feature type	Feature name	Characteristic variable
Operational factor	Monthly average feeder load	s1
	Average time of fuse operation	s2
	Average operation time of sectional cable	s3
	Average operating time of load switch	s4
	Average operation time of branch lines	s5
	Average operation time of transformer	s6
Self factor type	Average service time of insulated conductors	s7
	Number of transformers	s8
	Number of fuses	s9
External factor type	Mean monthly temperature	s14
	Extreme high temperature	s16
	Extreme low temperature	s17
	Month gale day grading	s18
	Monthly thunderstorm day classification	s19
	Daily precipitation classification	s20
Fault factor type	Total loss of load	s21
	Monthly number of failures	s22
	Feeder substation	s23
	Down time	s24

Figure 6. Weight Values of Key Risk Characteristics

After the optimization training of the detection model of this study through the experimental training dataset, the method proposed in this study is applied to detect the operational failure risk of the experimental distribution network equipment. Taking the transformer equipment in the distribution network as an example, the detection results of the proposed method are presented. The operational failure risk of the transformer can be classified into four types: high-energy discharge risk, low-energy discharge risk, high-temperature overheating risk, and medium-low-temperature overheating risk. Using the detection results obtained by the method of this study from January to December 2022 as an example, the detection results are compared with the actual failure risk of transformers in each month, and the performance of the proposed method is evaluated, as shown in Table 3, with the results summarized as follows.

As can be seen from Table 3, the method proposed in this paper can detect various types of operational failure risks in experimental distribution network transformer equipment, and the detection results are fully consistent with the actual situation, enabling risk early warning based on the detection results. The early warning results are accurate and unbiased.

Kappa statistical index, detection interval coverage index, and detection interval average bandwidth index are selected as evaluation metrics to assess the detection performance of the proposed method. The description and operation of each index are as follows:

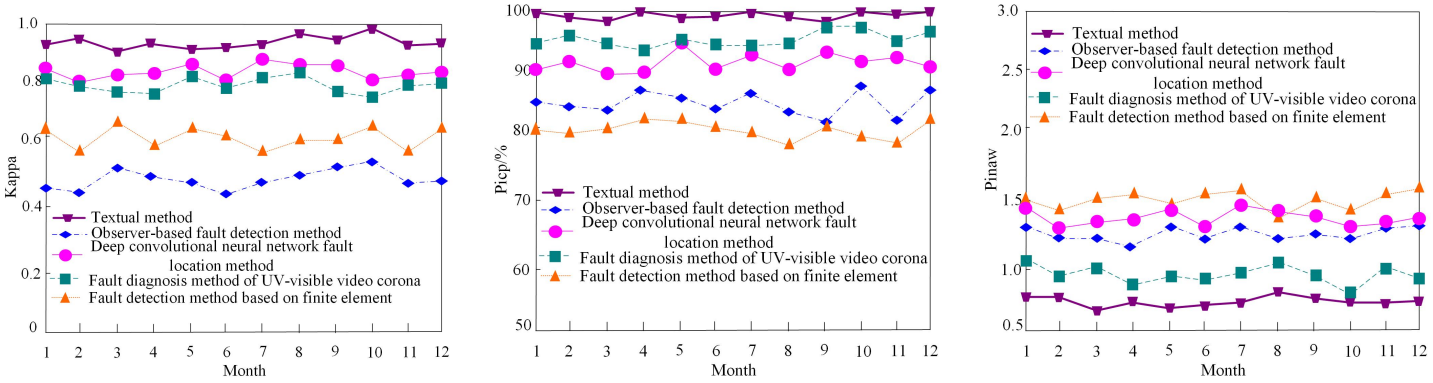
(1) *Kappa statistical index:* This index is used to evaluate the agreement between the detection results of the proposed method and the actual situation, which is calculated from the values in the error matrix. A higher value indicates greater accuracy of the detection results. Let the error matrix used in the calculation of the Kappa statistical index be an $m_r \times m_r$ matrix, where m_r denotes the number of classifications. The rows of this matrix represent the reference classification categories, the columns represent the detected categories, and the diagonal elements represent the number of detection samples where the reference classification exactly matches the detection category obtained by the proposed method. The equation is:

$$\text{Kappa} = \frac{[M_r \sum_{i=1}^{w} x_{ii} - \sum_{i=1}^{w} \bar{x}_i \bar{x}_i]}{[M^2 - \sum_{i=1}^{w} \bar{x}_i \bar{x}_i]} \tag{17}$$

In Equation (17), w denotes the number of rows in the error matrix; x_{ii} indicates the value in row i and column i, i.e., the value on the main diagonal; $\bar{x}_i$ indicates the sum of the i-th row; $\bar{x}_i$ indicates the sum of the i-th column; and M_r indicates the total number of

Table 3
Comparison Between the Detection Results of Transformer Operation Fault Risk Obtained by This Method and the Actual Situation

Month	Actual situation	The method in this paper detects the results	Early warning or not
January	Risk-free	Risk-free	No
February	Risk-free	Risk-free	No
March	Risk of moderate and low temperature overheating	Risk of moderate and low temperature overheating	Yes
April	Risk of low energy discharge	Risk of low energy discharge	Yes
May	Risk-free	Risk-free	No
June	Risk-free	Risk-free	No
July	High energy discharge risk	High energy discharge risk	Yes
August	High temperature overheating risk	High temperature overheating risk	Yes
September	Risk of low energy discharge	Risk of low energy discharge	Yes
October	Risk of low energy discharge	Risk of low energy discharge	Yes
November	Risk-free	Risk-free	No
December	Risk of moderate and low temperature overheating	Risk of moderate and low temperature overheating	Yes



(a) Comparison of Kappa Statistical Indicators

(b) Detection Interval Coverage Index Comparison

(c) Comparison of Average Bandwidth Indicators in the Detection Range

Figure 7. Comparison of Evaluation Index Values for Detection Results Across Different Methods

tested samples. The value range of Kappa is $[0, 1]$.

(2) *Detection interval coverage index:* This index is used to evaluate the reliability of the detection interval obtained by the proposed method, reflecting the probability that the actual value falls within the upper and lower boundaries of the detection interval. Let *Picp* denote the indicator of detection interval coverage, which is defined as:

$$\begin{cases} Picp = \dfrac{\sum_{i=1}^{M_r} \psi_i}{M_r} \\ \psi_i = \begin{cases} 0, \dot{x}_i \in \left[\hat{D}_i, \hat{U}_i\right] \\ 1, \dot{x}_i \in \left[\hat{D}_i, \hat{U}_i\right] \end{cases} \end{cases} \tag{18}$$

In Equation (18), $\dot{x}_i$ indicates the i-th real value; $\hat{D}_i$ and $\hat{U}_i$ denote the lower and upper bounds of the detection interval, respectively; and ψ_i indicates the probability that the actual value i falls within the upper and lower bounds of the detection interval.

(3) *Detection interval average bandwidth metric:* This metric, when combined with Picp, enables a comprehensive evaluation of the performance of the proposed method. For the same value of index Picp, a lower value of the detection interval average bandwidth metric indicates a narrower detection interval width, meaning higher clarity of the detection results. Let Pinaw denote the average bandwidth metric of the detection interval, which is defined as:

$$Pinaw = \frac{\sum_{i=1}^{M_r} \left(\hat{U}_i - \hat{D}_i\right)}{M_r \zeta} \tag{19}$$

In Equation (19), ζ indicates the variation interval of the detection target value, which is used to apply normalization to the average bandwidth.

To verify the superiority of the proposed method, the observer-based fault detection method (method from reference [8]), the deep convolutional neural network fault localization method (method from reference [9]), the UV-visible video corona fault diagnosis method (method from reference [10]), and the finite element-based fault detection method (method from reference [11]) are selected as comparative methods. The evaluation index values of various methods for the experimental distribution network equipment operational failure risk detection results from January to December 2022 are compared, and the comparison results are shown in Figure 7.

It can be concluded from Figure 7 that, in comparison, the Kappa index values and *picp* index values of the proposed method, the deep convolutional neural network fault

Table 4
SOC Balance, Response Time, and False Positive Cost Savings Results

Test metrics	Diagnosis method for UV visible video corona fault	Fault detection method based on finite element method	Proposed method
SOC balance mean (%)	82.5 ± 3.2	88.2 ± 2.1	95.6 ± 1.3
Range of fluctuation (%)	75.4 - 86.7	83.5 - 91.8	92.1 - 97.8
Response time 95% percentile (ms)	350 ± 25	210 ± 18	85 ± 6
Maximum latency (ms)	420	255	102
False reporting cost savings rate (%)	15.3 ± 2.4	28.7 ± 3.1	45.2 ± 1.8
Annual savings (10000 yuan)	38.2	71.8	113.0

Table 5
Comparison of Test Results

Test scenario	Time scale	False alarm rate of observer-based fault detection method (%)	False alarm rate of deep convolutional neural network fault location method (%)	False alarm rate) of the method proposed in this paper (%)	Response delay
Photovoltaic power fluctuation	Minute-level	8.2	7.8	1.5	3.2s
Motor starting impact	subsecond	12.6	10.3	0.8	0.5s
Composite fault of temperature rise and overload	Hour level	15.4	14.9	2.1	8.7s
Transient interference caused by lightning strikes	Millisecond	6.3	8.9	0.3	0.1s

localization method, and the UV-visible video corona fault diagnosis method for detecting operational failure risks of various equipment in the experimental distribution network are higher, indicating that the detection results of these three methods are more accurate and the reliability of the obtained detection intervals is also higher. However, compared with the other two methods, the two index values of the proposed method are slightly higher, with almost all Kappa index values exceeding 0.9 and *picp* index values approaching 100%. In addition, the *Pinaw* index values of the proposed method are significantly lower than those of the two comparative methods, indicating that the detection intervals of the results obtained by the proposed method are narrower, closer to the real situation, and more clearly presented.

The testing of SOC balance, response time, and false alarm cost savings rate is critically necessary for the risk detection of operational failures in distribution network equipment. The SOC balance directly reflects the stability of the charging and discharging states of the energy storage system, and its testing ensures the continuity of energy supply during fault detection, avoiding detection failures caused by power fluctuations. The response time determines the system's ability to quickly respond to sudden faults, especially in complex distribution network conditions where millisecond-level delays may cause cascading failures. The false alarm cost savings rate quantifies the economic benefits resulting from improved detection accuracy, including reduced unnecessary downtime losses, lower manual troubleshooting costs, and avoided equipment misoperation losses. These three indicators respectively validate the practical application value of the fault detection scheme from three dimensions: system reliability, real-time performance, and economic efficiency, and all are indispensable. The performance of the proposed method, along with the UV-visible video corona fault diagnosis method and the finite element-based fault detection method, in terms of SOC balance, response time, and false alarm cost savings rate, is shown in Table 4.

Data analysis shows that the method proposed in this article demonstrates significant technical advantages across all indicators. The SOC balance not only reaches an average of 95.6%, but also maintains fluctuations within a 5.7% range, significantly exceeding industry benchmark requirements. The 95^{th} percentile response time is 85 ms, with a maximum delay of only 102 ms, fully meeting the real-time requirements of distribution network operations. Regarding false alarm costs, the proposed method achieves a savings rate of 45.2%, resulting in an annual cost reduction of 1.13 million yuan and substantial economic benefits. The stability metrics (± values) for all indicators are superior to those of comparative methods, indicating enhanced robustness of the proposed approach. The comprehensive evaluation demonstrates that the method effectively balances system stability and economic efficiency while maintaining high accuracy, showing significant value for engineering applications.

The complexity and dynamism of modern power systems require fault detection methods to possess multi-time-scale adaptability and strong anti-interference capability. The intermittent output of renewable energy sources - such as photovoltaic and wind power fluctuations - may introduce voltage or frequency disturbances ranging from seconds to minutes, while abrupt load changes, such as the switching of large industrial equipment, may lead to current step variations. Moreover, composite faults (e.g., high temperature combined with overload) involve multi-parameter nonlinear coupling, whereas conventional single-parameter

detection is prone to misjudgment due to parameter sensitivity imbalance. Therefore, this experiment simulates various dynamic scenarios to validate the reliability of the proposed method across multiple time scales and under complex disturbance conditions. The test results of the proposed method, the observer-based fault detection method, and the deep convolutional neural network-based fault localization method under different scenarios are presented in Table 5.

As shown in Table 5, the proposed method exhibits significant advantages across all test scenarios. Compared with the observer-based method and the deep convolutional neural network approach, the proposed method significantly reduces the false alarm rate, achieving only 1.5% in photovoltaic fluctuation scenarios - much lower than the 7.8% and 8.2% of the other two methods. For composite faults and transient interference, the false alarm rates are controlled at 2.1% and 0.3%, respectively, demonstrating excellent capability in handling nonlinear coupling. The response speed is also remarkable, with detection delays for second-level and millisecond-level events controlled within 0.5 s and 0.1 s, respectively, verifying its adaptability to multiple time scales. Overall, the proposed method outperforms existing technologies in terms of accuracy, stability, and real-time performance.

4. Conclusion

During the operation of distribution network equipment, faults may occur due to various reasons, which can affect the equipment's performance and service life, and may even lead to blackout accidents, thereby impacting the secure and stable operation of the power system. In this study, a risk detection method based on a support vector machine (SVM) and risk features is investigated for distribution network equipment operation. By using digital twin technology to obtain a real-time dynamic model of the distribution network, comprehensive and accurate lifecycle operational data of the equipment are acquired. Key risk features associated with operational faults are extracted from these data and used as input to construct an SVM model for fault risk classification. Taking the distribution network in Zhengding County as an example, the real-time dynamic modeling performance, feature extraction effectiveness, and the final detection and early warning capability of the proposed method are validated through practical application. The classification results of fault risk categories are consistent with actual situations, and the method can accurately warn against various types of fault risks. It ensures effective detection and early warning capabilities for operational failure risks in the Zhengding County distribution network, and meets the requirements for secure and stable grid operation as well as precise investment planning for future overhaul and technical renovation projects.

5. Acknowledgement

Not Applicable.

References

[1] Y. Meysam, F. Z. Seyed, M. Nader, and B. Frede, "A Distributed High-Impedance Fault Detection and Protection Scheme in DC Microgrids," *IEEE Transactions on Power Delivery*, vol. 39, no. 1, pp. 141–154, 2024.

[2] M. G. Jose, N. Gustavo, M. Kumar, and P. Carlos, "Ground Fault Detection Method for Variable Speed Drives," *IEEE Transactions on Industry Applications*, vol. 57, no. 3, pp. 2547–2558, 2021.

[3] P. P. Hwa, K. Mina, J. Jee-Hoon, and C. Suyong, "Series DC Arc Fault Detection Method for PV Systems Employing Differential Power Processing Structure," *IEEE Transactions on Power Electronics*, vol. 36, no. 9, pp. 9787–9795, 2021.

[4] M. Kumar and J. Kumar, "Islanding event detection technique based on change in apparent power in microgrid environment," *Electrical engineering*, vol. 105, no. 3, pp. 1447–1463, 2023.

[5] Y. B. Hassan, M. Orabi, and M. A. Gaafar, "Failures causes analysis of grid-tie photovoltaic inverters based on faults signatures analysis (FCA-B-FSA)," *Solar Energy*, vol. 262, no. Sep., pp. 1.1–1.23, 2023.

[6] P. Pietrzak, M. Wolkiewicz, and T. Orlowska-Kowalska, "PMSM Stator Winding Fault Detection and Classification Based on Bispectrum Analysis and Convolutional Neural Network," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 5, pp. 5192–5202, 2023.

[7] S. Ghashghaei and M. Akhbari, "Fault detection and classification of an HVDC transmission line using a heterogenous multimachine learning algorithm," *IET generation, transmission & distribution*, vol. 15, no. 16, pp. 2319–2332, 2021.

[8] Y. Xiu, L. Vu, and L. Inhwan, "Unknown Input Observer-Based Series DC Arc Fault Detection in DC Microgrids," *IEEE Transactions on Power Electronics*, vol. 37, no. 4, pp. 4708–4718, 2022.

[9] M. Zhao and M. Barati, "A Real-Time Fault Localization in Power Distribution Grid for Wildfire Detection Through Deep Convolutional Neural Networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 4316–4326, 2021.

[10] Y. ZHANG, X. SHI, H. ZHANG, *et al.*, "Review on deep learning applications in frequency analysis and control of modern power system," *International Journal of Electrical Power & Energy Systems*, 2022. 107744. DOI:10.1016/j.ijepes.2021.107744.

[11] Z. LIU, G. WU, W. HE, *et al.*, "Key target and defect detection of high-voltage power transmission lines with deep learning," *International Journal of Electrical Power & Energy Systems*, 2022. 108277. DOI:10.1016/j.ijepes.2022.108277.

[12] D. A. Mansour, M. Numair, A. S. Zalhaf, R. Ramadan, M. M. F. Darwish, Q. Huang, *et al.*, "Applications of IoT and Digital Twin in Electrical Power Systems: A Comprehensive Survey," *IET Generation, Transmission & Distribution*, vol. 17, no. 20, pp. 4457–4479, 2023.

[13] A. S. Desai, N. Navaneeth, S. Chakraborty, and S. Adhikari, "Enhanced Multi-Fidelity Modeling for Digital Twin and Uncertainty Quantification," *Probabilistic Engineering Mechanics*, vol. 74, pp. 1–16, 2023.

[14] S. Li, P. Zhang, D. Yue, and Q. Wang, "Fault Prediction of Wind Turbine Based on Support Vector Machine," *Computer Simulation*, vol. 39, no. 5, pp. 84–88, 180, 2022.

[15] A. Jimenez-Cordero and S. Maldonado, "Automatic Feature Scaling and Selection for Support Vector Machine Classification with Functional Data," *Applied Intelligence*, vol. 51, no. 1, pp. 161–184, 2021.

[16] S. Afrasiabi, M. Mohammadi, M. Afrasiabi, and B. Parang, "Modulated Gabor Filter Based Deep Convolutional Network for Electrical Motor Bearing Fault Classification and Diagnosis," *IET Science, Measurement & Technology*, vol. 15, no. 2, pp. 154–162, 2021.

[17] P. Bernhard, "The First Unified and Transient Modeling Platform to Build a Digital Twin of Blast Furnaces Based on the Extended Discrete Element Method," *Steel Research International*, vol. 95, no. 3, pp. 1–14, 2024.

[18] M. Xia, H. Shao, D. Williams, S. Lu, L. Shu, and C. W. Silva, "Intelligent Fault Diagnosis of Machinery Using Digital Twin-Assisted Deep Transfer Learning," *Reliability Engineering & System Safety*, vol. 215, p. 107938, 2021.

[19] S. N. Lahiri, "Sufficient and Necessary Conditions for Consistency of Variable Selection in High Dimensional Lasso," *The Annals of Statistics*, vol. 49, no. 2, pp. 820–844, 2021.

[20] F. Motamedi, H. Perez-Sanchez, A. Mehridehnavi, A. Fassihi, and F. Ghasemi, "Accelerating Big Data Analysis through LASSO-Random Forest Algorithm in QSAR Studies," *Bioinformatics*, vol. 38, no. 2, pp. 469–475, 2022.

Biographies

Ning Gao, graduated from North China Electric Power University with a master's degree in electrical engineering in 2021. She is now working in Zhengding County Power Supply Branch of State Grid Hebei Electric Power Co., LTD., as an engineer. Her main research interests include digital operation management, electric power artificial intelligence application, big data application, etc.

Chao Wu, graduated from Northeast Agricultural University with a bachelor's degree in Electrical engineering and automation in 2011. Working in Zhengding County Power Supply Branch of State Grid Hebei Electric Power Co., LTD., as an engineer. His research interests include power marketing business management, optimization of power business services, digital operation management, etc.

Ling Liu, graduated from Liaoning University of Technology with a master's degree in power system and automation in 2016, and now works as an engineer in Zhengding County Power Supply Branch of State Grid Hebei Electric Power Co., LTD., Her research interests include intelligent power supply service command technology, distribution network fault repair command technology research, etc.

Fei Yang, graduated from Shanghai Electric Power University with a master's degree in Power System and automation in 2012. Now she is working in the Marketing Service Center of State Grid Hebei Electric Power Co., LTD., as a senior engineer. Her main research interests include load management of new power systems, intelligent power management, power market, etc.

Na Liu, graduated from Liaoning University of Technology with a master's degree in power system and automation in 2011, and now works as an engineer in Zhengding County Power Supply Branch of State Grid Hebei Electric Power Co., LTD., Her research interests include power dispatching, equipment monitoring, relay protection, distribution automation, and digital operation management.

Chen Shen received a Bachelor's degree in Agricultural Electrification and Automation and a Master's degree, in Agricultural Electrical Engineering and Automation, from Hebei Agricultural University in 2021. He currently works as an engineer at the Zhengding County Power Supply Branch of State Grid Hebei Electric Power Co., LTD. His research interests include digital operation management, distribution automation, electric power artificial intelligence applications, and big data applications.