IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 31 May 2024, accepted 6 July 2024, date of publication 10 July 2024, date of current version 2 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3426269

RESEARCH ARTICLE

A Transmission Line Fault Diagnosis Model Based on Interpretable BRB With Power Set

BOYING ZHAO[1], KANGLE LI[2], BING XU[3], GAIXIA GE[4], AND WEI HE[1]
[1]School of Computer Science and Information Engineering, Harbin Normal University, Harbin 150025, China
[2]School of Computer and mathematics, Harbin Finance University, Harbin 150030, China
[3]School of Economics and Management, Harbin Normal University, Harbin 150025, China
[4]Donghu Road School, Harbin 150079, China

Corresponding author: Bing Xu (xvbing@hrbnu.edu.cn)

This work was supported in part by the Postdoctoral Science Foundation of China under Grant 2020M683736, in part by the Teaching Reform Project of Higher Education in Heilongjiang under Grant SJGY20210456, in part by the Natural Science Foundation of Heilongjiang Province of China under Grant LH2021F038, in part by the Social Science Foundation of Heilongjiang Province of China under Grant 21GLC189, and in part by the Graduate Innovation Project of Harbin Normal University under Grant HSDSSCX2023-3.

ABSTRACT It is crucial to diagnose transmission line faults quickly and accurately. Belief rule base (BRB) has strong nonlinear modeling ability, allowing for the effective utilization of both qualitative knowledge and quantitative data related to these faults. To ensure accurate diagnosis results, the model must account for uncertainty and ignorance. Additionally, interpretability of the results is essential for improving diagnostic credibility. Therefore, a transmission line fault diagnosis model based on interpretable BRB with power set (PBRB-I) is proposed in this paper. Firstly, transmission line faults and data are analyzed, with the use of the Spearman correlation coefficient for data preprocessing. Secondly, according to the characteristics of transmission lines, interpretable modeling criteria are defined. Then, a power set identification framework is utilized to represent ignorance. Finally, the evidence reasoning (ER) algorithm is applied as a reasoning tool, and a parameter optimization method with interpretable constraints based on the projection covariance matrix adaptive evolutionary strategy (P-CMA-ES) is proposed. In the case study, the PBRB-I model demonstrates an accuracy of 91.11%, and it exhibits high performance stability across different data allocation ratios. It not only shows outstanding accuracy but also effectively expresses ignorance and produces interpretable results.

INDEX TERMS Belief rule base, power set, interpretability, fault diagnosis, transmission line.

I. INTRODUCTION

Transmission lines are vital components in power systems [1], facilitating efficient and reliable transmission of electric energy across long distances [2]. The expansion of transmission lines in different terrain and geographical locations easily results in faults [3], causing potential disruptions to power supply. In severe cases, transmission line faults can lead to major power accidents such as power grid disconnection [4] and collapse [5]. Therefore, it is essential to diagnose faults in transmission lines with both speed and accuracy.

Fault diagnosis of transmission lines has always been a hot topic for scholars worldwide. At present, the main methods of transmission line fault diagnosis include physical models, data-driven models and hybrid models [6]. The physical model establishes the corresponding model from the essential characteristics. Wang et al. proposed a single-ended fault location approach that utilized a distributed parameter model, which was specifically designed for half-wavelength lines. The method is unaffected by fault location and type, and has high location accuracy [7]. Chuncheng and Jiao gave a prediction model of the icing quality and thickness of transmission line conductors with uniform icing and non-uniform icing by introducing the collision coefficient, freezing coefficient and collection coefficient [8]. However, understanding the system's operating principles is fundamental to constructing a physical model. A physical model requires a detailed description of the system's physical behavior. This increases the difficulty of building and applying physical models.

The data-driven model mines the characteristic information of system faults from a large amount of data by analyzing the data. Guo et al. developed a stacked-informer network to extract the hidden features contained in long-sequence

The associate editor coordinating the review of this manuscript and approving it for publication was Sarasij Das.

time series data and combined gradient concentration (GC) technology with an optimizer [9]. Shakiba et al. proposed a convolutional neural network-based diagnostic system that analyzes factors such as phase difference between connecting buses and fault resistance. This analysis demonstrates the robustness of the method. The method can detect and estimate fault locations between fault zones [10]. Theoretically, data-driven models have the potential to improve accuracy through training, but they also have certain application limitations. Data-driven models require a large amount of high-quality training data to achieve high accuracy. However, in practical applications, especially for some rare types of faults, it may be difficult to obtain sufficient sample data, which can directly affect the reliability of the diagnostic results [11].

The hybrid model is a synthesis of the above two models. Hou et al. predicted transmission line damage probability by simulating random wind fields using the extreme value type I probability distribution and the Monte Carlo method [12]. Teimourzadeh et al. used the cownvolutional neural network (CNN) and the hybrid model of deep reinforcement learning (DRL) to detect the operation fault of transmission lines [3]. The hybrid model combines the advantages of physical models and data-driven models, providing a more powerful and flexible diagnostic tool [28]. It not only improves diagnostic accuracy and reliability but also addresses the limitations of using a single model in applications. By leveraging the theoretical foundation of physical models and the high accuracy of data-driven models, hybrid models can offer more reliable and effective fault diagnosis solutions for complex application scenarios.

In the actual fault diagnosis of transmission lines, there are two problems that need to be considered. First, considering that transmission lines are affected by many disturbing factors [13], such as equipment state changes and natural meteorological disasters, it is necessary to adopt rigorous technology with the ability to deal with uncertain information for fault diagnosis, which reduces the influence of uncertain factors on diagnosis accuracy. Second, because data-driven models require a substantial amount of data and complex algorithms for model training, the modeling process is often opaque [14], which can be challenging for individuals to comprehend intuitively and explain the diagnosis results of the model. However, interpretability is of great significance in fault diagnosis of transmission lines, which can support the reliability of the model.

In this complex scene, the belief rule base (BRB) can show powerful adaptability and can be regarded as an excellent model choice. First, BRB has outstanding nonlinear modeling ability, which is capable of effectively handling uncertain information, including fuzziness, uncertainty, inconsistency and randomness [15]. It can comprehensively apply expert knowledge and limited data samples and maintain good performance even with small training sample sizes [16]. Second, BRB is a modeling method based on If-Then rules [17] and has an easy-to-understand knowledge expression process. BRB has been applied in many fields, such as safety assessment [29] and behavior prediction [30].

Nonetheless, employing BRB for transmission line fault diagnosis faces two problems. First, the improvement of diagnosis accuracy often becomes the theme of model optimization, but interpretability is ignored, which affects the credibility of the model [18]. Second, through the actual diagnosis, it has been discovered that there are similarities between different types of fault data, which results in local ignorance and then reduces the accuracy of diagnosis. The power set effectively addresses ignorance by encompassing both the signal sets and all of their subsets [19]. This comprehensive approach ensures that a wide range of possibilities and potential variations are considered. Therefore, a transmission line fault diagnosis model based on an interpretable BRB with a power set (PBRB-I) is proposed in this paper. Next, the reasoning procedure utilizes the evidential reasoning (ER) algorithm, and an optimization algorithm with interpretable constraints based on the projection covariance matrix adaptive evolution strategy (P-CMA-ES) is designed according to interpretable modeling criteria defined by the characteristics of transmission lines.

This paper makes the following key contributions:

1) A PBRB-I model is proposed to implement the fault diagnosis of transmission lines. By preprocessing the data, the complexity of the model is reduced.

2) Using the power set identification framework to address the issue of ignorance, by comprehensively considering all possible combinations, avoids missing any potential diagnostic results, thereby ensuring the reliability of the diagnosis.

3) An optimization algorithm with interpretability constraints is designed. It enables the model to not only have efficient diagnostic capabilities but also provide clear explanations and decision-making rationales.

The remaining sections of this paper are organized as follows: In Section II, the formulation of problems related to transmission line fault diagnosis is presented, and the construction process is outlined. Section III describes the structure, reasoning and optimization. In Section IV, the superiority of the PBRB-I model is demonstrated via a case study. Section V concludes this paper.

II. PROBLEM DESCRIPTION

Section II-A describes four problems in transmission line fault diagnosis. Section II-B describes the construction of the PBRB-I model.

A. PROBLEM DESCRIPTION

To build an interpretable BRB with a power set for transmission line fault diagnosis, the following four problems must be addressed.

1) The redundancy among various fault features affects the diagnosis results. Therefore, it is necessary to select a set of feature subsets that can clearly distinguish fault types through

an evaluation criterion to improve the accuracy of diagnosis.

$$X^* = g(X, \Omega) \tag{1}$$

where X represents original features of faults, $X^* = \{X_1, \cdots, X_M\}$ represents selected features, $g(\cdot)$ represents the evaluation criterion for feature selection, Ω represents the parameter set in the selection.

2) Interpretability helps to increase the credibility of fault diagnosis results, so to guarantee the model's interpretability, the basic requirements to be met in the modeling process should be defined as follows:

$$\text{principle:}\{d|d_1, \cdots, d_n\} \tag{2}$$

where d represents the n interpretability criteria.

3) According to the interpretability criteria, the diagnosis process is defined as follows:

$$y = f(X^*, \Phi, d) \tag{3}$$

where Φ is the parameter set of the diagnosis procedure for transmission lines, $f(\cdot)$ is the diagnosis process, and y is the diagnostic output result.

4) To enhance the diagnostic accuracy, the model's parameters should be optimized under interpretable criteria.

$$\Phi_{best} = optimize(X^*, y, O, d) \tag{4}$$

where $optimize(\cdot)$ is the optimization function, Φ_{best} is the parameter set after optimization, and O is the parameter set of the optimization.

B. CONSTRUCTION OF THE MODEL

To address the problems summarized in Section II-A, the PBRB-I model is proposed. Its output can be defined as follows:

$$Type = \{D_1, D_2, \ldots, D_N\} \tag{5}$$

where $D_1, \ldots, D_N$ is N fault types of the transmission line. In its fault diagnosis, global ignorance refers to a scenario where the fault could be any one of all N fault types, and local ignorance refers to a scenario where the fault could be any J types of all N fault types, where $J < N$. The set of fault types with ignorance is described as follows [20]:

$$2^{Type} = \{\emptyset, D_1, D_2, \ldots, D_N, \{D_1, D_N\}, \ldots, \{D_1, \ldots, D_{N-1}\}, Type\} \tag{6}$$

where $\emptyset$ is an empty set, $\{D_1, D_N\}$ represents that the fault diagnosis result of the transmission line may be D_1 or D_N, which is employed to illustrate local ignorance, and $Type$ is used to describe global description. According to the above definition, 2^{Type} fault probabilities can be obtained.

The BRB consists of belief rules, and the kth rule of the PBRB-I model can be described as follows:

$$\begin{aligned}
&\text{IF} x_1 \text{ is } A_1^k \wedge x_2 \text{is } A_2^k \wedge \cdots \wedge x_N \text{is } A_N^k \\
&\text{THEN } y \text{ is } \{(D_1, \beta_{1,k}), (D_2, \beta_{2,k}), \cdots, (D_{2^{Type}}, \beta_{2^{Type},k})\}, \\
&\sum_{m=1}^{2^{Type}} \beta_{m,k} = 1 \\
&\text{with } \textit{rule weight } \theta_k, k \in \{1, 2, \ldots, R\} \\
&\text{and } \textit{attribute weight } \delta_1, \delta_2, \ldots, \delta_i, i \in \{1, 2, \ldots, N\} \\
&\text{in } d_1, d_2, \ldots, d_n
\end{aligned} \tag{7}$$

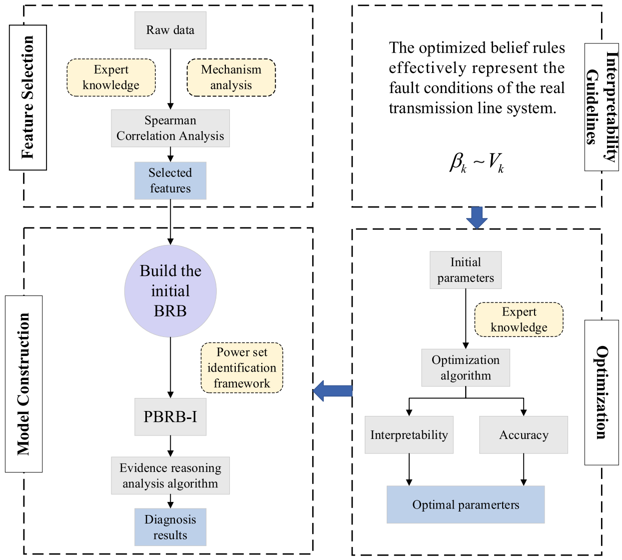

FIGURE 1. The modeling process of PBRB-I.

where $x_1, \ldots, x_N$ represents antecedent attributes of the fault diagnosis model for transmission lines, $A_1^k, \ldots, A_N^k$ represents reference values corresponding to the N attributes, $D_1, \ldots, D_{2^{Type}}$ represents the diagnosis results, $\beta_{1,k}, \ldots, \beta_{2^{Type},k}$ represents belief degrees corresponding to the 2^{Type} diagnosis results, θ_i represents the rule weight of the kth rule among R rules, and $\delta_1, \ldots, \delta_i$ represents the i attribute weight of the kth rule.

III. TRANSMISSION LINE FAULT DIAGNOSIS MODEL

The modeling procedure for the PBRB-I model is detailed in this section. Section III-A defines the basic structure of the transmission line fault diagnosis model based on the PBRB-I. Section III-B defines the model's interpretability. Section III-C describes the reasoning procedure. Section III-D describes the optimization.

The PBRB-I fault diagnosis model for transmission lines, as shown in Fig.1, includes the following steps:

1) Through Spearman correlation analysis, highly correlated fault features can be selected, significantly reducing dimensionality, and improving the accuracy and interpretability of the results.

2) Initial construction of the PBRB-I model. The power set identification framework is applied to describe global ignorance and local ignorance.

3) Reasoning on the model, and utilizing the P-CMA-ES algorithm with interpretability constraints to optimize the fault diagnosis model, to enhance interpretability and accuracy.

A. THE BASIC STRUCTURE OF THE PROPOSED MODEL

The basic structure of the model encompasses the following steps:

1) DEFINE THE INPUT ATTRIBUTES OF THE MODEL

Feature selection plays an important role in optimizing model performance [21]. By selecting typical features from the raw data, the dimension of the input data is reduced, and so are unnecessary calculations. At the same time, the interpretability of the model is improved, and it is easier for people to understand the implementation process of diagnosis.

Spearman correlation analysis is a feature selection method suitable for nonlinear data that can make full use of the correlation between variables when selecting the best features, thus reducing unnecessary features and improving the effect of the model [22]. Consequently, Spearman correlation analysis is employed for data preprocessing in this paper.

The Spearman correlation coefficient possesses a value range of -1 to 1, where a value above 0 denotes a positive correlation, and a value below 0 signifies a negative correlation. As the absolute value of the coefficient approaches 1, the correlation strength increases. The Spearman correlation coefficient is calculated as follows [23]:

$$\rho = \frac{\sum_{i=1}^{c}(R_i - \bar{R})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^{c}(R_i - \bar{R})^2 \sum_{i=1}^{c}(T_i - \bar{T})^2}} \tag{8}$$

where R and T represent two features, $\bar{R}$ and $\bar{T}$ represent the averages of the two features, and c represents the number of samples of each feature.

2) DEFINE THE OUTPUT

The output is defined according to the fault types of the transmission line. The transmission line fault types are described as follows:

$$Type = \{D_1, D_2, D_3, D_4, D_5, D_6\} \tag{9}$$

The BRB discriminative framework is ineffective in representing the local ignorance information that arises from similar attributes, while the power set can describe ignorance well [24]. Therefore, this paper uses an identification framework with a power set, which is explained as follows:

$$\begin{aligned} 2^{Type} = \{&\emptyset, D_1, \ldots, D_6, \\ &\{D_1, D_2\}, \ldots, \{D_5, D_6\} \\ &\{D_1, D_2, D_3\}, \ldots, \{D_4, D_5, D_6\} \\ &\{D_1, \ldots, D_4\}, \ldots, \{D_3, \ldots, D_6\} \\ &\{D_1, \ldots, D_5\}, \{D_2, \ldots, D_6\} \\ &Type\} \end{aligned} \tag{10}$$

3) GENERATE BELIEF RULES

Setting the belief distribution according to expert knowledge not only combines the mechanism of transmission lines but is also more practical and beneficial to enhancing the reliability of the results. The kth rule is described as follows:

$$\begin{aligned} &\text{IF } x_1 \text{ is } \mathrm{A}_1^k \wedge x_2 \text{ is } \mathrm{A}_2^k \wedge x_3 \text{ is } \mathrm{A}_3^k \\ &\quad \text{THEN } y \text{ is } \{(\emptyset, \beta_{1,k}), (D_1, \beta_{2,k}), \ldots, (D_6, \beta_{7,k}) \\ &\quad (\{D_1, D_2, \}\beta_{8,k}), \ldots, (\{D_5, D_6\}\, \beta_{22,k}), \\ &\quad (\{D_1, D_2, D_3\}\, \beta_{23,k}), \ldots, (\{D_4, D_5, D_6\}, \beta_{42,k}), \\ &\quad (\{D_1, \ldots, D_4, \}\beta_{43,k}), \ldots, (\{D_3, \ldots, D_6\}, \beta_{57,k}), \\ &\quad (\{D_1, \ldots, D_5, \}\beta_{58,k}), \ldots, (\{D_2, \ldots, D_6\}, \beta_{63,k}), \\ &\quad (Type, \beta_{64,k})\}, \\ &\quad \text{with } \theta_k \\ &\quad \text{and } \delta_1, \delta_2, \delta_3 \\ &\quad \text{in } d_1, d_2, \ldots, d_n \end{aligned} \tag{11}$$

B. INTERPRETABILITY OF THE MODEL

Criteria 8: Optimized belief rules need to meet the requirements of the actual system. As the main interpretable aspect of BRB, belief rules can transform expert knowledge into models and provide accurate semantic descriptions of the relationships between inputs and outputs [25]. Reasonable forms of belief distribution should reflect the true relationship between inputs and outputs, conform to the expectations of the actual system, and help users understand the reasoning process of the system. However, people often focus on improving accuracy while neglecting the loss of interpretability. For example, reasonable belief distributions should be monotonic or convex, as shown in Fig.3(a). However, concave belief distributions, as shown in Fig.3(b), are unreasonable and may occur when accuracy is overly emphasized. This distribution not only violates the logic of the actual system but may also lead to unnecessary confusion. Therefore, meeting this guideline is crucial.

Only when belief rules meet the requirements of the actual system and reasonably reflect the relationship between inputs and outputs can the BRB system truly leverage its advantages in decision support and knowledge transmission, providing effective support for practical applications.

To make the belief distribution conform to the mechanism and reality, the constraints are as follows:

$$\begin{aligned} &\beta_k \sim V_k (k = 1, \ldots, L) \\ &\quad V_k \in \{\{\beta_1 \le \beta_2 \le \cdots \le \beta_{2^{Type}}\} \\ &\quad or\{\beta_1 \ge \beta_2 \ge \cdots \ge \beta_{2^{Type}}\} \\ &\quad or\{\beta_1 \le \cdots \le \max(\beta_1, \beta_2, \cdots, \beta_{2^{Type}}) \ge \cdots \ge \beta_{2^{Type}}\}\} \end{aligned} \tag{12}$$

where V_k represents the interpretability constraint in the kth rule.

C. THE REASONING PROCEDURE OF THE MODEL

Drawing upon the preceding analysis, the fault features applied to the model are selected, and the fault diagnosis model based on the PBRB-I model is constructed. The

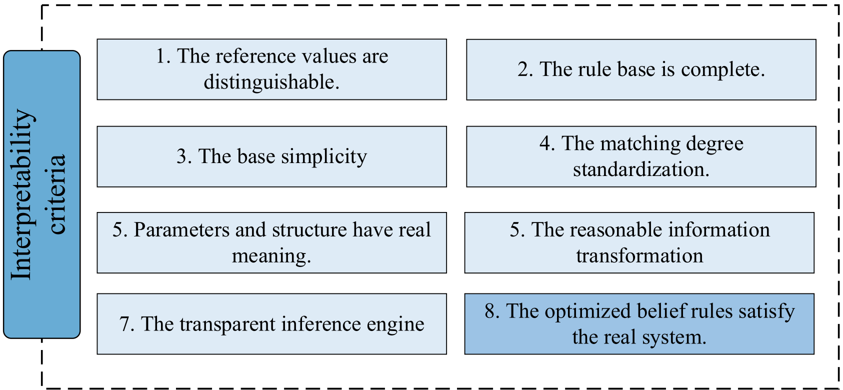

FIGURE 2. General interpretability criteria.

reasoning procedure based on the ER algorithm is summarized as follows:

Step 1: Calculate the matching degree.

Through the calculation of the following formula, the matching degree a_i^k between the ith input and the kth rule can be calculated.

$$a_i^k = \begin{cases} \dfrac{A_i^{l+1} - x_i}{A_i^{l+1} - A_i^l} & k = l, A_i^l \le x_i \le A_i^{l+1} \\ 1 - a_i^k & k = l+1 \\ 0 & k = 1 \cdots K, k \ne l, l+1 \end{cases} \tag{13}$$

where A_i^k is the kth reference value of the ith input.

Step 2: Calculate the activation weight.

The activation weight w_k of the kth rule is obtained through this equation:

$$w_k = \frac{\theta_k \prod_{i=1}^{N} (a_i^k)^{\delta_i}}{\sum_{i=1}^{K} \theta_l \prod_{i=1}^{N} (a_i^l)^{\delta_i}} \tag{14}$$

Step 3: Generate the belief degree.

The belief degree and utility value of the nth diagnosis result are calculated below:

$$\beta_n = \frac{\mu \times [\prod_{i=1}^{L} (w_l \beta_{n,l} + 1 - w_l \sum_{i=1}^{N} \beta_{i,l}) - \prod_{i=1}^{L} (1 - w_l \sum_{i=1}^{N} \beta_{i,l})]}{1 - \mu \times [\prod_{i=1}^{L} (1 - w_l)]} \tag{15}$$

$$\mu = \frac{1}{\sum_{n=1}^{N} \prod_{l=1}^{L} (w_l \beta_{n,l} + 1 - w_l \sum_{i=1}^{N} \beta_{i,l}) - (N-1) \prod_{l=1}^{L} (1 - w_l \sum_{i=1}^{N} \beta_{i,l})} \tag{16}$$

Step 4: Calculate the final output.

$$\mu(S(A')) = \sum_{n=1}^{N} \mu(D_n)\beta_n \tag{17}$$

where $\mu(D_n)$ is the utility value of D_n, A' is output vectors in real systems, $S(\cdot)$ is a set of belief distributions, and $\mu(S(A'))$ is the final expected utility value. The final belief distribution y is represented as follows:

$$y = \{(D_n, \beta_n), n = 1, \ldots, N\} \tag{18}$$

D. OPTIMIZATION OF THE PROPOSED MODEL

The P-CMA-ES algorithm is an effective method for practical optimization, with several advantages: 1) The ability to handle large-scale optimization problems. 2) The ability to converge rapidly towards the global optimal solution. 3) Its high robustness. Therefore, utilizing the P-CMA-ES algorithm is indeed a favorable optimization choice, especially for situations requiring the management of large-scale optimization problems and aiming to attain efficient and high-quality solutions.

Incorporating an optimization algorithm can greatly enhance the accuracy of model diagnosis. However, to enhance the accuracy, the optimization algorithm may choose some specific working methods and parameters, which makes the realization process of the model complicated and difficult to understand, so with the improvement of accuracy, the interpretability will decrease. Because the fault diagnosis of transmission lines is very critical, it is also crucial to consider the interpretability of the model while ensuring accuracy. Therefore, this paper proposes a modified P-CMA-ES optimization algorithm with interpretable constraints, and the objective optimization function of the PBRB-I model is defined as:

$$\begin{aligned} & \min MSE(\Phi) \\ & s.t. \sum_{n=1}^{N} \beta_{k,n} = 1 \\ & \quad 0 \le \theta_k \le 1 \quad k = 1, \ldots, R \\ & \quad 0 \le \beta_{k,n} \le 1 \quad n = 1, \ldots, 2^M \end{aligned}$$

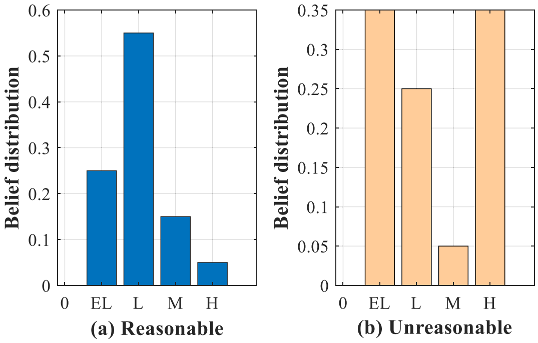

FIGURE 3. Reasonable and unreasonable belief distribution.

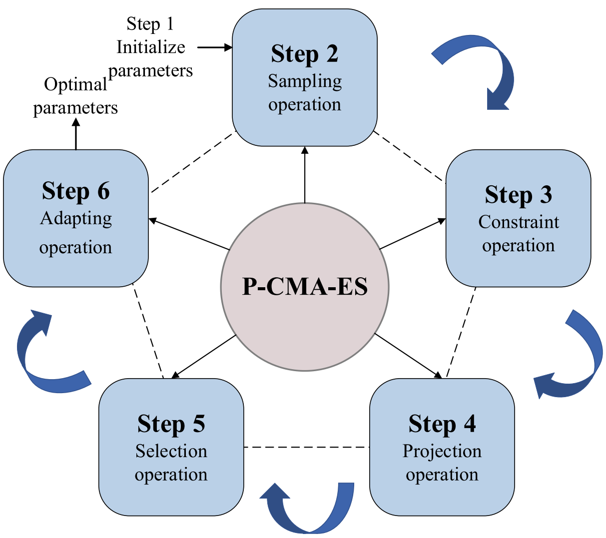

FIGURE 4. Optimization process of the modified P-CMA-ES algorithm.

$$\begin{aligned} &0 \leq \delta_i \leq 1 \quad i = 1, \ldots, N \\ &\beta_k \sim V_k \end{aligned} \tag{19}$$

where $MSE(\cdot)$ is the error between the diagnosis result and the real situation, which can be expressed as:

$$MSE = \frac{1}{t}\sum_{i=1}^{t}(y_i - \tilde{y}_i)^2 \tag{20}$$

where t is the total number of training samples, y_i is the real value of the transmission line, and $\tilde{y}_i$ is the fault diagnosis result of the PBRB-I model.

The steps of the modified P-CMA-ES optimization algorithm with interpretable constraints are shown in Fig. 4, as follows:

Step 1: Initialize parameters

The parameters to be initialized include the following: Φ, the initial average value ave, the number of solutions in the population P, the number of solutions in the optimal subgroup S, the dimension of the problem d, the optimal subgroup e, and the weight of the optimal subgroup h_i. Φ was initialized as follows:

$$\begin{aligned} \Phi = [&\theta_1, \ldots, \theta_R, \beta_{1,1}, \ldots, \beta_{1,2^M}, \\ &\ldots, \beta_{R,1}, \ldots, \beta_{R,2^M}, \delta_1, \ldots, \delta_N] \end{aligned} \tag{21}$$

Step 2: Sampling operation

$$\Phi_v^{t+1} = ave^t + \gamma^t G(0, C^t), v = 1, \ldots, h \tag{22}$$

where Φ_v^{t+1} is the vth solution vector generated for the $(t+1)$th time, γ is the step size, G is the normal distribution, and C^t is the covariance matrix of the tth generation population.

Step 3: Constraint operation

Adjust unreasonable rules according to the eighth interpretable criterion.

$$\beta_k^{t+1} \sim Criterion_8, k = 1, 2, \ldots, d \tag{23}$$

where β_k^{t+1} stands for the newly generated belief distribution that satisfies the eighth interpretability criterion $Criterion_8$.

Step 4: Projection operation

This process makes the solution vector satisfy the constraints of optimization, which is summarized as follows:

$$\begin{aligned} &\Phi_i^{t+1}(1 + b \times (\varepsilon - 1) : b \times \varepsilon) \\ &= \Phi_i^{t+1}(1 + b \times (\varepsilon - 1) : b \times \varepsilon) - L^T \times (L \times L^T)^{-1} \\ &\times \Phi_i^{t+1}(1 + b \times (\varepsilon - 1) : b \times \varepsilon) \times L \end{aligned} \tag{24}$$

where the number of variables b satisfies $b = (1 \ldots B)$, and B is the solutions of each equality constraint. a, when the constraints are equal, the number is expressed by ε. In addition, $L = [1, 1, \ldots, 1]_{1\times R}$ is used to express a parameter vector.

Step 5: Selection operation

The optimal parameter solutions that meet the conditions are selected, and the mean value and covariance matrix of the population are updated accordingly.

$$ave^{t+1} = \sum_{i=1}^{S} q_i \Phi_i^{t+i}, \sum_{i=1}^{S} q_i = 1 \tag{25}$$

where q_i is the weight of the ith solution in the optimal subgroup,$1 \leq i \leq S$.

Step 6: Adapting operation

The method for updating the covariance matrix is as follows:

$$\begin{aligned} C^{t+1} &= (1 - le_1 - le_S)T^t + le_1 s_c^{t+1}(s_c^{t+1})^T \\ &+ le_S \sum_{i=1}^{S} t_i\left(\frac{\Phi_i^{t+1} - ave^t}{\gamma^t}\right) \times \left(\frac{\Phi_i^{t+1} - ave^t}{\gamma^t}\right)^T \end{aligned} \tag{26}$$

$$s_c^{t+1} = (1 - le_c)s_c^t + \sqrt{le_c(2 - le_c)(\sum_{i=1}^{S} t_i^2)^{-1}} \times \frac{ave^t ave^{t+1}}{\varepsilon^t} \tag{27}$$

$$\varepsilon^{t+1} = \varepsilon^t \exp\left(\frac{e_\varepsilon}{O_\varepsilon}\left(\frac{||s_c^{g+1}||}{||H(0, I_m)||} - 1\right)\right) \tag{28}$$

$$\begin{aligned} s_\varepsilon^{t+1} &= (1 - le_\varepsilon)s_\varepsilon^t + \sqrt{le_c(2 - le_c)(\sum_{i=1}^{S} t_i^2)^{-1}} \\ &\times C^{t-\frac{1}{2}} \times \frac{ave^{t+1} - ave^t}{\varepsilon^t} \end{aligned} \tag{29}$$

where le_1, le_S, le_c, le_ε represent learning rates, s_ε^t represents the tth evolutionary step, and $s_\varepsilon^t = 0$. In addition, I_m represents the identity matrix, o_η represents the damping coefficient, and $H(0, C^t)$ represents the mathematical expectation.

IV. CASE STUDY

Section IV-A introduces the dataset used in this case study. Section IV-B introduces the processing of raw data. In Section IV-C, the initial PBRB-I model is constructed. In Section IV-D, the interpretability of the PBRB-I model is verified. In Section IV-E, compared with the diagnosis results of other methods, the results show the effectiveness of this model.

A. DATASET DESCRIPTION

The three-phase transmission line system fault classification data set from literature [27] is used in this paper. A, B, and C represent the three phases of the transmission line system, and G represents ground. The input of the model is three-phase voltage and three-phase current, which are denoted as la, lb, lc, Va, Vb and Vc respectively. The faults in this dataset include six states, including normal (N) and five fault states. The faults are LG faults (between phase A and ground), LL faults (between phase A and phase B), LLG faults (between phase A and phase B and ground), LL faults (between all three phases) and LLLG faults (three phase symmetric faults).

B. DATA PREPROCESSING

In this paper, Spearman correlation analysis is used as the method of feature selection, and the relatively useful features for the target variable are selected for modeling, which provides assistance to enhance the accuracy and reliability and reduce the complexity.

Through the comprehensive consideration of correlation and significance, the weak features are removed, so lb, lc and va are input into the model as attributes.

Controlling the number of input attributes is crucial for reducing model complexity. By selecting the most relevant attributes through correlation analysis, not only can computational and storage costs be reduced, but the interpretability and generalization ability of the model can also be improved. In this chapter, if all six attributes are input into the model, the model's complexity significantly increases. Assuming each attribute is assigned four reference values, 4^6rules would be generated, leading to a combinatorial explosion of rules. However, after correlation analysis, the number of rules is reduced to 4^3 rules, a reduction of 98.44% compared to the former.

C. THE INITIAL MODEL CONSTRUCTION

After many practical verifications and mechanism analyses, expert knowledge with rich experience and theory has been formed. The model founded on expert knowledge can better adapt to the complex actual situation of transmission lines and has high interpretability [25].

The core algorithm of PBRB-I model proposed in this paper is based on Matlab. The setting of model reference points is through the analysis of data and the mechanism of transmission line faults. In Table 1, the three attributes are assigned four semantic values in turn, including extremely low (EL), low (L), middle (M), and high (H). The fault types of the transmission line are shown in Equation (30), and six reference values are used to indicate the fault types in Table 2.

TABLE 1. Attribute reference points.

Attribute	Attribute weight	EL	L	M	H
lb	1	-875	-55	191	884
lc	1	-875	15	173	845
va	1	-0.3	0.04	0.3	0.8

TABLE 2. Output reference points.

Reference points	N	LG	LLG	LL	LLL	LLLG
Reference value	0	1	2	3	4	5

$$Type = \{N, LG, LLG, LL, LLL, LLLG\} \tag{30}$$

The basic structure of the model is demonstrated, and the power set identification framework based on fault types is introduced in Section III-A. By analyzing the data of faults, it is concluded that there is only local ignorance between adjacent faults in *Type*. Therefore, the power set identification framework for transmission line fault diagnosis is expressed as follows:

$$\begin{aligned} Type^* = \{&N, \{N, LG\}, LG, \{LG, LLG\}, LLG, \\ &\{LLG, LL\}, LL, \{LL, LLL\}, LLL, \{LLL, LLG\}\} \end{aligned} \tag{31}$$

D. VERIFICATION OF THE INTERPRETABILITY OF THE MODEL

Expert knowledge can guide model optimization. The optimization of the PBRB-I model is a process of gradually approximating expert knowledge. However, the BRB model lacks the ability to approach expert knowledge during the optimization process, leading to an overemphasis on accuracy and a decline in interpretability. Fig. 5 compares the belief distributions of expert knowledge, BRB, and PBRB-I, showing how different models retain expert knowledge features during optimization. When the optimized belief distribution more closely resembles expert knowledge, more key features of expert knowledge are preserved. The belief distribution optimized by PBRB-I better fits expert knowledge. For example, in rules 1, 4, 5, and 10, PBRB-I accurately describes the actual faults of transmission lines. The PBRB-I model proposed in this paper considers interpretability criteria during optimization, preserving more features of

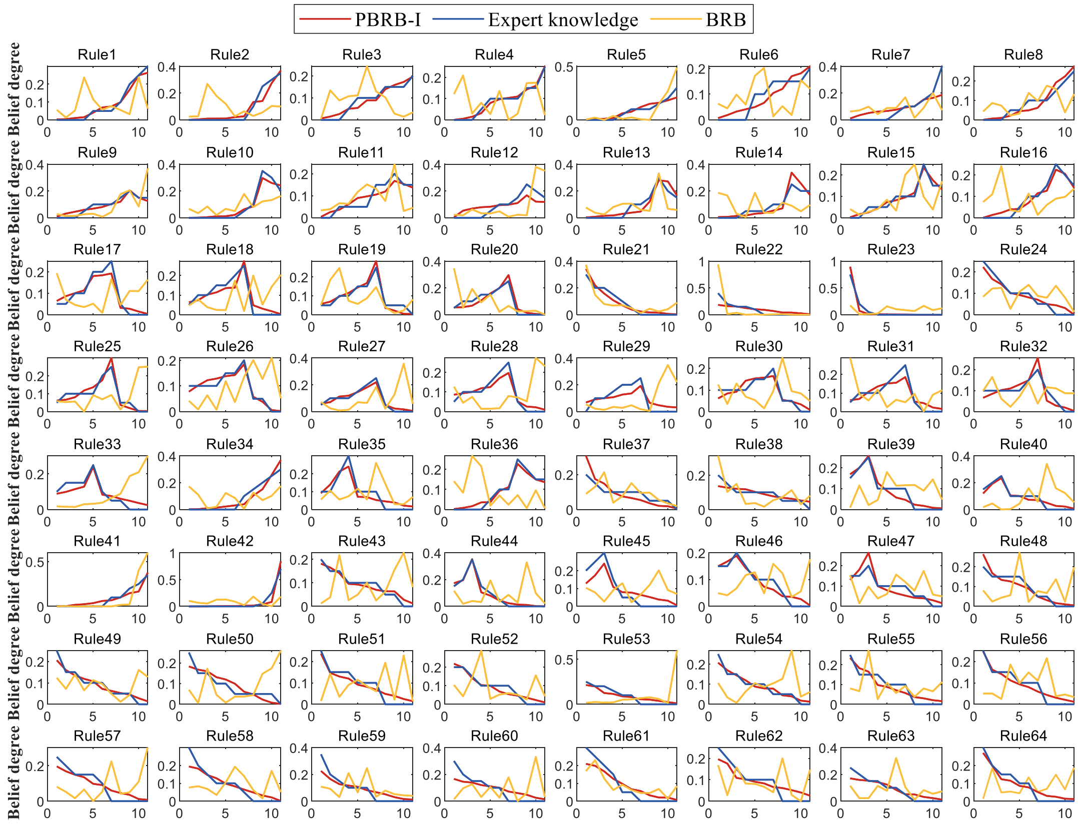

FIGURE 5. The belief distribution comparison.

expert knowledge, thus having stronger interpretability and credibility. However, the BRB model produces many rules that do not conform to the transmission line mechanism during optimization, such as rules 2 and 16, leading to compromised interpretability. This phenomenon indicates that in random model optimization, the preservation of expert knowledge is neglected, thereby affecting the overall performance of the model.

Fig. 6 depicts the belief distribution of different diagnosis results of PBRB-I model. It is evident that PBRB-I model provides a clear semantic description of transmission line faults.

E. VERIFICATION OF THE MODEL

To validate the effect of the proposed PBRB-I model on the fault diagnosis of transmission line faults, the following evaluation indicators are introduced in this section:

1) Overall accuracy. The overall accuracy O_acc refers to the diagnostic accuracy for all test samples All in the diagnostic task. This indicator can comprehensively and intuitively describe the overall performance of the model, and the calculation method is as follows:

$$O\_acc = \frac{TN}{All} \times 100 \tag{32}$$

where TN represents the number of samples correctly diagnosed.

2) Fault diagnosis accuracy. The fault diagnosis accuracy F_acc considers the diagnostic accuracy of fault samples FN, which is targeted and can be obtained by the following methods:

$$F\_acc = \frac{FN^*}{FN} \times 100 \tag{33}$$

where FN^* represents the number of correctly diagnosed fault samples.

3) Fault detection rate. The fault detection rate f_rate refers to the model's accuracy in fault detection, focusing on the model's ability to distinguish between the normal state and fault state. The methods are as follows:

$$f\_rate = \frac{FN'}{FN} \times 100 \tag{34}$$

where FN' indicates the total number of detected faults.

Through the above three evaluation indexes, the diagnosis results obtained from the experiment are analyzed. The overall accuracy reflects the correct classification ability of the model on the whole test set, and the fault diagnosis accuracy reflects the ability of PBRB-I model to judge transmission line faults. The fault detection rate shows that PBRB-I model can find faults of transmission lines in time, which is beneficial to the maintenance and overhaul of transmission lines.

The data are randomly divided into six groups: 8:2, 7:3, 6:4, 5:5, 4:6 and 3:7. The model undergoes training under

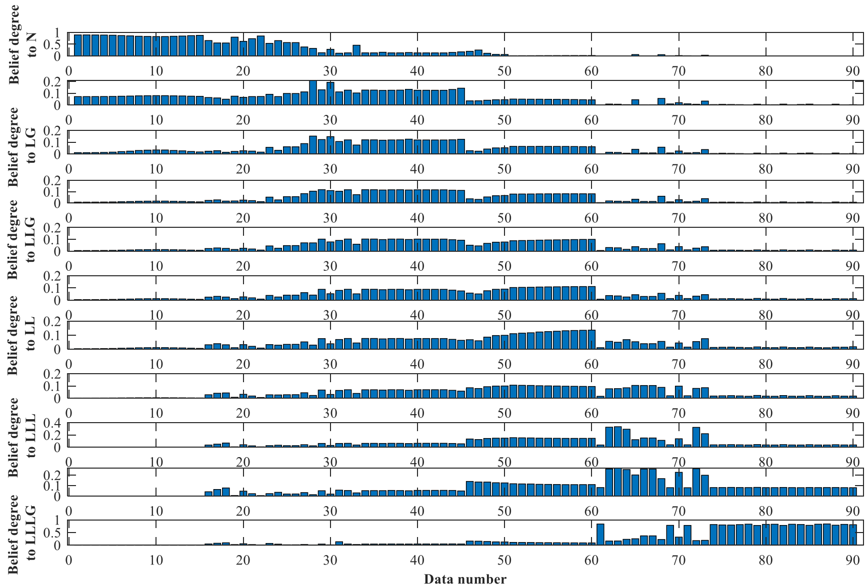

FIGURE 6. The belief distributional results generated by PBRB-I.

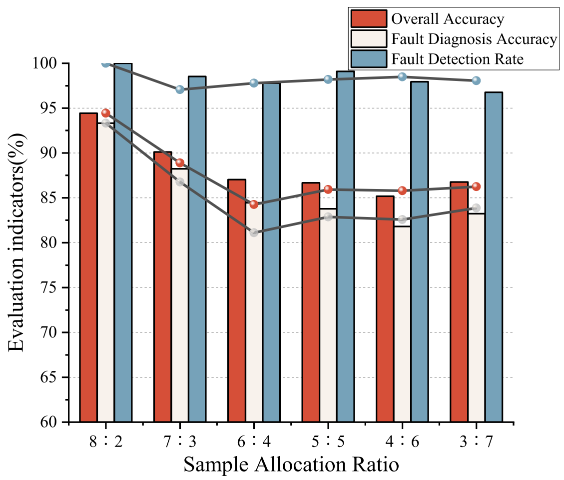

FIGURE 7. Comparison of diagnostic results under different sample ratios.

each training set, following which it is tested on the corresponding test set. The experimental results under different sample ratios show that the PBRB-I model maintains high accuracy and the index value fluctuates little, as shown in Fig. 7. In Fig. 7, the column represents the three indexes of the PBRB-I model, and the broken line reflects the three indexes of the BRB model. This indicates that these two models still show good performance even when provided with limited samples and possess the capability to train small samples. Meanwhile, the PBRB-I model exceeds the BRB model in accuracy. Therefore, the PBRB-I model can provide higher reliability and accuracy for small sample fault diagnosis.

Mean square error (MSE) is also introduced as an auxiliary indicator in this section. According to Fig. 8, with the increase of iteration times, the accuracy increases and the MSE decreases, indicating that the model gradually approaches the best diagnosis effect. When the model has been trained for 900 times, the diagnosis effect tends to be stable, showing good diagnosis ability.

TABLE 3. Comparative experiments.

Method	Accuracy
PBRB-I	91.11%
BPNN	81.11%
KNN	82.2%
LDA	72.2%
Decision Tree	79.4%

F. COMPARATIVE EXPERIMENTS

In this section, the back propagation neural network (BPNN), the k-Nearest Neighbor (KNN), the Latent Dirichlet Allocation (LDA) and the Decision Tree are used as comparative experiments to validate the effectiveness of the PBRB-I model. The effects of different models verified based on the training set and the test set are shown in Table 3. The accuracy of PBRB-I model for fault diagnosis of transmission lines exceeds the above four methods. In addition to its superior accuracy, PBRB-I model also has the following advantages:

1) The PBRB-I model combines transmission line mechanisms and expert knowledge. The mechanisms provide a scientific theoretical foundation, ensuring the reliability of the model. The rule-based modeling approach makes it easy to understand. Through clear rules, users can easily comprehend the diagnostic logic of the model.

2) The reasoning process of the PBRB-I model is transparent. In practical applications, this transparency helps

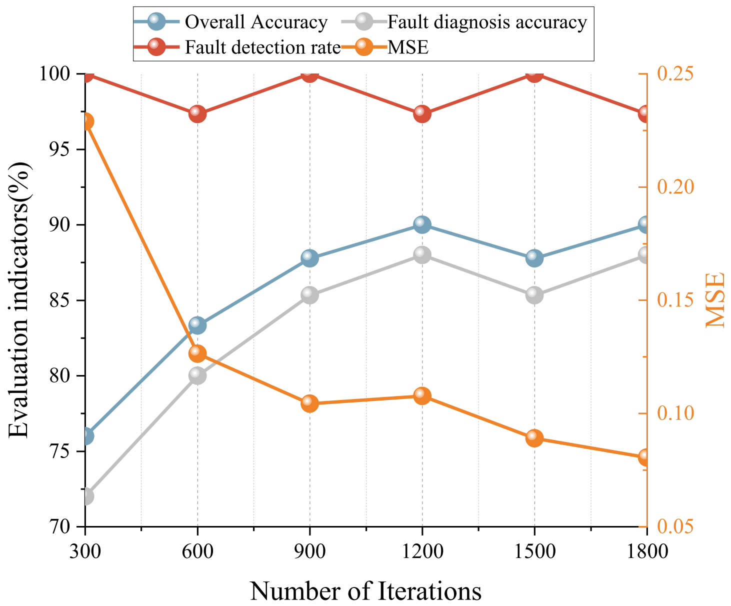

FIGURE 8. Performance comparison under different iterations.

verify the rationality and accuracy of the process, thereby ensuring the reliability of the final results.

3) The PBRB-I model is interpretable, clearly explaining how inputs are mapped to outputs. The belief distributions generated by the PBRB-I model during diagnosis align with real-world situations, making the model more reliable and practical in actual applications.

V. CONCLUSION

A transmission line fault diagnosis model based on the PBRB-I is proposed in this paper, which solves four problems: the treatment of uncertain information, the transparency of the modeling process, the protection of interpretability in the optimization process and the ability to deal with ignorance. First, the original data are preprocessed by feature selection to reduce the complexity of the model. Second, the power set identification framework is employed to solve the problem of local ignorance, and the initial model is constructed according to expert knowledge. Third, because interpretability is often ignored in the optimization, an improved P-CMA-ES algorithm with interpretable constraints is designed.

The PBRB-I model retains the advantages of the BRB model while better describing local ignorance in transmission line fault diagnosis, effectively handling uncertain information. This comprehensive diagnostic approach can significantly improve the accuracy and reliability of the diagnosis. Additionally, by utilizing the improved P-CMA-ES optimization algorithm with interpretable constraints designed in this study, the initial BRB model has been optimized, further improving the model's accuracy without compromising its interpretability. Interpretability allows the steps of fault diagnosis to be clearly understood and tracked. For critical equipment like transmission lines, transparency in the diagnostic process is crucial, helping technicians and decision-makers understand and trust the diagnostic results.

Numerous factors affect the faults of large equipment such as transmission lines, and considering more attribute inputs aids in comprehensive fault analysis and diagnosis. Therefore, it is worthwhile to consider incorporating more attribute inputs. Thus, comprehensively considering more fault attributes has significant research value. Interpretability is of great significance to the healthy management of transmission lines, and how to keep the interpretability of the PBRB-I model in structural optimization deserves further study.

ACKNOWLEDGMENT

(Boying Zhao and Kangle Li contributed equally to this work.)

REFERENCES

[1] B. Bajwa, C. Butani, and C. Patel, "A novel approach towards predicting faults in power systems using machine learning," *Electr. Eng.*, vol. 104, no. 1, pp. 363–368, Feb. 2022.
[2] J. Yu and X. Zhang, "A cross-layer wireless sensor network energy-efficient communication protocol for real-time monitoring of the long-distance electric transmission lines," *J. Sensors*, vol. 2015, pp. 1–13, Jun. 2015.
[3] H. Teimourzadeh, A. Moradzadeh, M. Shoaran, B. Mohammadi-Ivatloo, and R. Razzaghi, "High impedance single-phase faults diagnosis in transmission lines via deep reinforcement learning of transfer functions," *IEEE Access*, vol. 9, pp. 15796–15809, 2021.
[4] L. Wang, W. Jiang, and W. Dong, "Design of an underground transmission line condition fault monitoring system for power grids based on data analysis algorithms," *J. Sensors*, vol. 2022, pp. 1–11, Oct. 2022.
[5] Y. Li, H. Li, S. Huang, J. Li, and H. Wu, "A local measurement-based phase sequence exchange emergency control approach," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 13828–13837, Dec. 2022.
[6] Z. Jizhong, L. Teng-yan, W. Wan-Li, L. Sheng-Lin, and D. Han-jiang, "Evaluation of operational reliability of integrated energy systems: Data driven method and model-data hybrid drive method," *Trans. China Electrotechnical Soc.*, vol. 37, no. 13, pp. 3227–3240, 2022.

[7] L. Wang, Y. Hu, X. Shi, X. Li, Y. Gong, and M. Deng, "A single-ended fault location method based on distributed parameter model for half-wavelength AC transmission line," in *Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2018, pp. 1–5.
[8] L. Chuncheng and L. Jiao, "Icing mechanism of transmission line wire and ice coating model of Ice Glaze," *High Voltage Technol.*, vol. 37, no. 1, pp. 241–248, 2011.
[9] L. Guo, R. Li, and B. Jiang, "A data-driven long time-series electrical line trip fault prediction method using an improved stacked-informer network," *Sensors*, vol. 21, no. 13, p. 4466, Jun. 2021.
[10] F. M. Shakiba, M. Shojaee, S. M. Azizi, and M. Zhou, "Real-time sensing and fault diagnosis for transmission lines," *Int. J. Netw. Dyn. Intell.*, vol. 1, pp. 36–47, Dec. 2022.
[11] C. Lefeng, Y. Tao, and Z. Xiaoshun, "Application and prospects of machine learning in the field of energy and power systems," *Autom. Electr. Power Syst.*, vol. 43, no. 1, pp. 15–31, 2019.
[12] H. Hou, S. Yu, H. Wang, Y. Xu, X. Xiao, Y. Huang, and X. Wu, "A hybrid prediction model for damage warning of power transmission line under typhoon disaster," *IEEE Access*, vol. 8, pp. 85038–85050, 2020.
[13] S. R. Fahim, S. M. Muyeen, M. A. Mannan, S. K. Sarker, S. K. Das, and N. Al-Emadi, "Uncertainty awareness in transmission line fault analysis: A deep learning based approach," *Appl. Soft Comput.*, vol. 128, Oct. 2022, Art. no. 109437.
[14] L. Himanen, A. Geurts, A. S. Foster, and P. Rinke, "Data-driven materials science: Status, challenges, and perspectives," *Adv. Sci.*, vol. 6, no. 21, Nov. 2019, Art. no. 1900808.
[15] Z.-J. Zhou, G.-Y. Hu, C.-H. Hu, C.-L. Wen, and L.-L. Chang, "A survey of belief rule-base expert system," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 51, no. 8, pp. 4944–4958, Aug. 2021.
[16] M. Chen, Z. Zhou, B. Zhang, G. Hu, and Y. Cao, "A novel combination belief rule base model for mechanical equipment fault diagnosis," *Chin. J. Aeronaut.*, vol. 35, no. 5, pp. 158–178, May 2022.
[17] Z. Feng, W. He, Z. Zhou, X. Ban, C. Hu, and X. Han, "A new safety assessment method based on belief rule base with attribute reliability," *IEEE/CAA J. Autom. Sinica*, vol. 8, no. 11, pp. 1774–1785, Nov. 2021.
[18] Y. Cao, Z. Zhou, C. Hu, W. He, and S. Tang, "On the interpretability of belief rule-based expert systems," *IEEE Trans. Fuzzy Syst.*, vol. 29, no. 11, pp. 3489–3503, Nov. 2021.
[19] Z. Zhou, Z. Feng, C. Hu, X. Han, Z. Zhou, and G. Li, "A hidden fault prediction model based on the belief rule base with power set and considering attribute reliability," *Sci. China Inf. Sci.*, vol. 62, no. 10, pp. 1–16, Oct. 2019.
[20] Z.-J. Zhou, G.-Y. Hu, B.-C. Zhang, C.-H. Hu, Z.-G. Zhou, and P.-L. Qiao, "A model for hidden behavior prediction of complex systems based on belief rule base and power set," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 48, no. 9, pp. 1649–1655, Sep. 2018.
[21] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.
[22] W. Weitong, F. Haidong, L. Chengsi, Z. Zhongyang, S. Yuhao, T. Chang, and Z. Chenghang, "Based on random forest algorithm Hedge boiler exit NO_x emissions prediction model research," *Thermal Power Gener.*, vol. 51, no. 4, pp. 96–104, 2022.
[23] N. Gui and J. Hua, "Data driven time delay identification and main steam temperature prediction in thermal power units," *J. Comput. Appl.*, vol. 40, no. 11, p. 3400, 2020.
[24] G.-W. Sun, W. He, H.-L. Zhu, Z.-J. Yang, Q.-Q. Mu, and Y.-H. Wang, "A wireless sensor network node fault diagnosis model based on belief rule base with power set," *Heliyon*, vol. 8, no. 10, Oct. 2022, Art. no. e10879.
[25] Z. Feng, R. Yang, Z. Zhou, and C. Hu, "Trustworthy fault diagnosis method based on belief rule base with multi-source uncertain information for vehicle," *IEEE Trans. Ind. Electron.*, vol. 71, no. 7, pp. 7947–7956, Jul. 2024.
[26] Y. Cao, Z. J. Zhou, C. H. Hu, S. W. Tang, and J. Wang, "A new approximate belief rule base expert system for complex system modelling," *Decis. Support Syst.*, vol. 150, Nov. 2021, Art. no. 113558.
[27] M. Jamil, S. K. Sharma, and R. Singh, "Fault detection and classification in electrical power transmission system using artificial neural network," *SpringerPlus*, vol. 4, no. 1, pp. 1–13, Dec. 2015.
[28] P. Han, B. Zhao, W. He, L. Kong, Y. Li, G. Zhou, and J. Feng, "An interpretable BRB model with interval optimization strategy for lithium battery capacity prediction," *Energy Sci. Eng.*, vol. 11, no. 6, pp. 1945–1959, Jun. 2023.
[29] X. Yin, W. He, Y. Cao, G. Zhou, and H. Li, "Interpretable belief rule base for safety state assessment with reverse causal inference," *Inf. Sci.*, vol. 651, Dec. 2023, Art. no. 119748.
[30] Q. Zhang, B. Zhao, W. He, H. Zhu, and G. Zhou, "A behavior prediction method for complex system based on belief rule base with structural adaptive," *Appl. Soft Comput.*, vol. 151, Jan. 2024, Art. no. 111118.

BOYING ZHAO received the B.Eng. degree from Qingdao Agricultural University, Qingdao, China, in 2022. She is currently pursuing the master's degree with Harbin Normal University, Harbin, China. Her research interests include belief rule base, deep learning, fault diagnosis, and interpretable diagnosis systems.

KANGLE LI received the master's degree from Northeastern University, in 2007. She is currently working with the School of Computer and Mathematics, Harbin Finance University. Her research interests include fault prediction and safety state assessment of large structural equipment.

BING XU received the Ph.D. degree in agricultural system engineering and management engineering from Northeast Agricultural University, Harbin, China, in 2018. She is currently working with the School of Economics and Management, Harbin Normal University, Harbin. Her research interests include belief rule base and agricultural system engineering and algorithm research.

GAIXIA GE received the bachelor's degree from Harbin Normal University. She is currently working with the Donghu Road School, Harbin. Her research interests include evidential reasoning and natural language processing.

WEI HE received the Ph.D. degree from Harbin University of Science and Technology, in 2018. He is currently working with the School of Computer Science and Information Engineering, Harbin Normal University. His research interests include evidential reasoning, deep learning, and belief rule base.