An Interpretable Bearing Fault Diagnosis Model Based on Hierarchical Belief Rule Base

Boying Zhao[1], Yuanyuan Qu[2], Mengliang Mu[2], Bing Xu[1*], and Wei He[1*]
[1] Harbin Normal University, Harbin 150025, China
[e-mail: zhaoboying0131@163.com (B.Z.);
xvbing@hrbnu.edu.cn (B.X.); hewei@hrbnu.edu.cn (W.H.)]
[2] Heilongjiang Agricultural Engineering Vocational College
Harbin 150088, China
[e-mail: candyuan@live.cn (Y.Q.); amusky@163.com (M.M.)]
[*]Corresponding author: Bing Xu and Wei He

Received August 21, 2023; revised January 21, 2024; accepted April 26, 2024; published May 31, 2024

Abstract

Bearings are one of the main components of mechanical equipment and one of the primary components prone to faults. Therefore, conducting fault diagnosis on bearings is a key issue in mechanical equipment research. Belief rule base (BRB) is essentially an expert system that effectively integrates qualitative and quantitative information, demonstrating excellent performance in fault diagnosis. However, class imbalance often occurs in the diagnosis task, which poses challenges to the diagnosis. Models with interpretability can enhance decision-makers' trust in the output results. However, the randomness in the optimization process can undermine interpretability, thereby reducing the level of trustworthiness in the results. Therefore, a hierarchical BRB model based on extreme gradient boosting (XGBoost) feature selection with interpretability (HFS-IBRB) is proposed in this paper. Utilizing a main BRB alongside multiple sub-BRBs allows for the conversion of a multi-classification challenge into several distinct binary classification tasks, thereby leading to enhanced accuracy. By incorporating interpretability constraints into the model, interpretability is effectively ensured. Finally, the case study of the actual dataset of bearing fault diagnosis demonstrates the ability of the HFS-IBRB model to perform accurate and interpretable diagnosis.

Keywords: belief rule base, bearing, fault diagnosis, extreme gradient boosting, class imbalance, interpretability

1. Introduction

Bearings, as critical components in rotating machinery [1], are widely used in fields such as aerospace, computer numerical control machine tools, and robotics. The stability of their operational condition directly affects the overall system's performance [2]. Therefore, to enhance the reliability of bearings, it is essential to conduct timely and reliable fault diagnosis on them, which has attracted extensive attention [3].

To date, various methods have been developed and employed for bearing fault diagnosis, primarily falling into three categories: model-driven methods, data-driven methods, and hybrid methods.

The model-driven bearing fault diagnosis method utilizes prior knowledge to construct a model for diagnosing faults. For example, Rzadkowski et al. used tip-timing and tip-clearance techniques to diagnose the faults of the middle bearing based on a numerical analysis of seventh stage blade free vibration [4]. The model-driven method typically requires less data for training and validating models. This advantage becomes particularly significant in situations where data are limited or obtaining large-scale data is challenging. However, employing model-driven methods demands the amalgamation of vast expertise from diverse domains, making the high-accuracy modeling of complex systems relatively difficult [5]. Additionally, the intricate structure of bearings and susceptibility of fault signals to noise interference pose a formidable challenge in the establishment of a diagnosis model based on the model-driven method, and any uncertainty may lead to false diagnosis [6].

The data-driven method predominantly relies on the collection, processing, and analysis of a large volume of actual data to learn features, patterns, and trends rather than depending on prior knowledge. For example, Fu et al. employed the ensemble empirical mode decomposition (EEMD) technique to extract essential features from data and harnessed the power of a support vector machine (SVM) to achieve accurate classification and prediction of these faults [7]. However, data-driven methods have a high requirement for data quality, and they may suffer from data bias and compromise model accuracy and generalization ability when faced with insufficient or imbalanced data samples. The modeling process needs to be more interpretable; otherwise, the results will be unconvincing [8].

The hybrid model effectively combines the strengths of both model-driven and data-driven methods. By integrating diverse models, it acquires valuable insights from data while incorporating prior knowledge. The method can guarantee interpretability and accuracy at the same time. As a result, the hybrid model excels in addressing a wide range of diverse problems and complex data distributions [9]. For example, belief rule base expert systems, fuzzy expert systems and fuzzy neural networks.

Belief rule base (BRB) is a rule-based modeling method proposed by Yang et al in 2006 [10]. It has demonstrated excellent performance and promising applications in various fields, such as fault diagnosis [11] and state assessment [12]. As a typical hybrid model, BRB exhibits excellent performance in handling numerical quantitative data and linguistic qualitative knowledge from diverse sources [13]. BRB possesses a strong capability for nonlinear modeling, enabling it to effectively represent complex causal relationships between antecedent attributes and consequents, even when confronted with different forms of uncertainty [14]. Moreover, BRB possesses interpretability, which gives it an advantage in explaining and understanding the model's decision-making process. This is especially important for application scenarios where results need to be interpreted and verified.

Bearings are commonly used in various industrial equipment and machinery, operating in complex and ever-changing environments. This diversity exposes the fault information to

various external factors, including temperature, humidity, load, vibration, etc., leading to uncertainty in the gathered information [15]. Meanwhile, it is essential to ensure the reliability of the diagnostic results. Interpretable fault diagnosis results can provide effective decision support. BRB provides an enhanced knowledge representation that accommodates both quantitative data and qualitative information with uncertainties. It excels in capturing complex nonlinear causal relationships between inputs and outputs, all the while maintaining a transparent reasoning process. Therefore, BRB can enhance the reliability and accuracy of diagnostic results, providing effective decision support and thus achieving efficient bearing fault diagnosis.

Nevertheless, there are two issues that require attention in the BRB-based bearing fault diagnosis model. Firstly, the class imbalance problem may result in significantly fewer samples in certain classes compared to others, leading to poorer classification performance on the minority classes. Despite the strong modeling and small-sample handling capabilities of the BRB, it can still be negatively affected by class imbalance. Secondly, to enhance the accuracy and generalization ability, optimizing the BRB with observed data is a requisite step. However, the stochastic nature of optimization algorithms can lead to different model results in different runs, thereby affecting the model interpretability. The disruption of interpretability hinders users from understanding the decision-making process, diminishing the reliability and acceptability of the model. Consequently, it limits the applicability and adoption in the corresponding domain. Therefore, maintaining a certain level of interpretability is crucial, especially for models involved in critical decision-making processes.

In response to the aforementioned two issues, an interpretable bearing fault diagnosis model with feature selection based on hierarchical BRB (HFS-IBRB) is proposed. Below are the contributions of this study:

(1) The utilization of a hierarchical structure in the BRB model addresses the challenges associated with multi-fault classes and class imbalance in bearing fault diagnosis.

(2) An interpretable optimization algorithm based on the projection covariance matrix adaptive evolution strategy (P-CMA-ES) algorithm is developed, which can maintain a high model performance while providing more interpretable diagnostic results.

This paper is organized as follows: In Section 2, specific problems to be addressed in bearing fault diagnosis using BRB are summarized. Section 3 elaborates on the construction, inference, and optimization of the HFS-IBRB model. In Section 4, the effectiveness of the model is validated through a case study. Finally, Section 5 presents a conclusion of the research content in this paper.

2. Problem formulation

Drawing from the limitations of BRB discussed in the preceding section, the HFS-IBRB model for bearing fault diagnosis is proposed. Constructing this model involves addressing the following challenges：

Problem 1: How to choose an appropriate model structure to address the issue of class imbalance is crucial for enhancing the robustness and diagnosis capability of rare faults in the model. Adopting a hierarchical model design is an effective approach to tackle class imbalance by introducing hierarchical relationships, increasing the focus on minority classes and improving the diagnosis capability of rare faults. Simultaneously, selecting suitable features contributes to increasing the sensitivity to rare classes and reducing the negative impact of irrelevant features on the model. The process of feature selection using the extreme gradient boosting (XGBoost) and the modeling procedure can be described as follows:

$$\{x_i^t, x_j^t\} = f(\{x_1, x_2, ..., x_m\}, \alpha) \tag{1}$$

$$y_t = g(x_i^t, x_j^t, C, \varphi) \tag{2}$$

where $f(\bullet)$ represents the process of analyzing the feature importance of the fault features using XGBoost. In this process, for each rule base output, the first two features that are most important to the output are selected from all fault features. This ensures that each rule base utilizes the most relevant features for further model reasoning. $\{x_1, x_2, ..., x_m\}$ represents all the bearing fault features, α represents the method used by XGBClassifier to calculate the fault feature importance, and x_i^t, x_j^t represent the selected features obtained through feature selection. In (2), $g(\bullet)$ represents the process of model construction. In the modeling process, a layered approach is adopted based on the independent feature selection strategy to better capture the complexity of the system, thereby enhancing the overall modeling effectiveness. y_t represents the diagnostic results of the t-th part of the model, C represents the set of interpretability criteria that need to be considered during the modeling process, and φ represents the model parameters.

Problem 2: By setting interpretability criteria, ensure the model interpretability. During each iteration of the optimization algorithm, the training process of the model introduces uncertainty, causing the model's behavior to vary slightly with each training, thereby reducing the interpretability. Therefore, the modeling process should meet the fundamental requirements for the interpretability of the system. By conducting a mechanistic analysis of the system, the defined requirements for interpretability in the modeling process can be described as the following set:

$$Interpretability: \{C \mid C_1, C_2, \cdots, C_a\} \tag{3}$$

where a represents the number of criteria, $C_1, C_2, \cdots, C_a$ represent the interpretability criteria.

Problem 3: How to design a reasonable optimization process is critical. However, the randomness in the optimization process often compromises the interpretability of the model. Therefore, considering the interpretability of the model, it is important to design a reasonable optimization process. This process is described as the following non-linear mapping relationship.

$$\varphi_{best} = optimize(x_i^t, x_j^t, y, \gamma, C) \tag{4}$$

where $optimize(\bullet)$ represents the optimization function, γ represents the parameter set in optimization, and φ_{best} represents the optimal parameters of the model. The interpretability criteria that incorporate into the optimization process can serve as guidelines for interpretability in the optimization process, which helps to prevent excessive pursuit of accuracy at the expense of interpretability.

3. Bearing fault diagnosis model based on the HFS-IBRB model

This paper proposes a bearing fault diagnosis model based on the HFS-IBRB to tackle the problems identified in Section 2 of the bearing fault diagnosis system. In Section 3.1, the diagnostic process of the HFS-IBRB is summarized. In Section 3.2, the model construction and feature selection method are explained. In Section 3.3, the interpretability is described. Sections 3.4 and 3.5 elaborate on the model inference and optimization in the diagnostic process sequentially.

3.1 The HFS-IBRB model fault diagnosis process

In "HFS-IBRB", "H" signifies the hierarchical structure, "FS" indicates feature selection for fault characteristics, and "I" stands for the interpretability possessed by the model.
The proposed fault diagnosis method for bearings is shown in **Fig. 1**. The adopted hierarchical structure fully considers the importance of both the entire dataset and the corresponding data for each fault category. In this hierarchical structure, different layers sequentially handle different aspects of the dataset, thereby synthesizing more accurate and interpretable results. To fully leverage the advantages of the hierarchical structure and ensure that each level considers the appropriate feature importance, it is necessary to partition the dataset [16], as shown in **Fig. 1**.

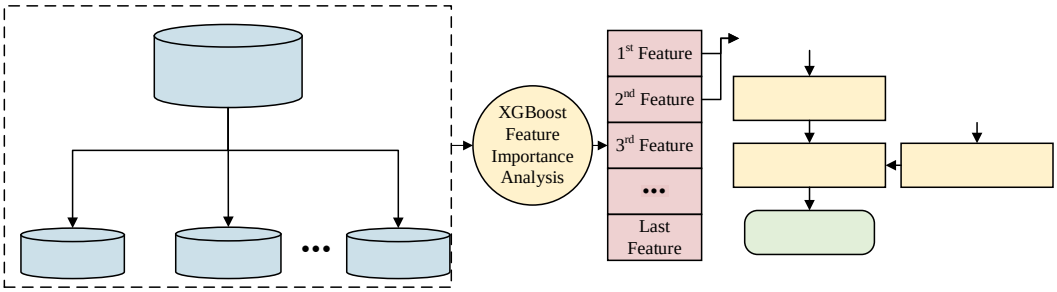

Fig. 1. Flow chart of the HFS-IBRB model diagnosis

3.2 Construction of the HFS-IBRB model

3.2.1 Model structure

BRB is composed of many belief rules [17], which provides an effective knowledge representation and inference framework for the model. For example, the k-th rule is as follows:

$$\begin{aligned}
&\text{IF } x_1 \text{ is } \mathrm{A}_1^k \wedge x_2 \text{ is } \mathrm{A}_2^k \wedge \cdots \wedge x_M \text{ is } \mathrm{A}_M^k \\
&\text{THEN } y \text{ is } \{(D_1,\beta_{1,k}),(D_2,\beta_{2,k}),\cdots,(D_N,\beta_{N,k})\}, \sum_{n=1}^{N}\beta_{n,k} \le 1 \\
&\text{with } \textit{rule weight } \theta_k, k \in \{1,2,...,R\} \\
&\text{and } \textit{attribute weights } \delta_i, i \in \{1,2,...,M\} \\
&\text{in } C_1, C_2, ..., C_a
\end{aligned} \tag{5}$$

where $x_1, x_2, ..., x_M$ represents antecedent attributes, $A_1^k, A_2^k, ..., A_M^k$ represents reference values corresponding to the M attributes, $D_1, ..., D_N$ represents the fault categories,

$\beta_{1,k},...,\beta_{N,k}$ represents belief degrees corresponding to the N categories, θ_k represents the rule weight, and δ_i represents the attribute weights.

The HFS-IBRB model consists of two layers. The first layer is the main BRB, responsible for delivering diagnostic outcomes across all fault categories. Assuming there are N categories, the main BRB can simultaneously handle N different fault categories. This design makes the main BRB a multi-class classifier capable of diagnosing all fault categories at the same time. Each sub-BRB will focus on two specific fault categories, and the diagnostic results from $N-1$ sub-BRBs cover all fault categories. The HFS-IBRB model structure is shown in **Fig. 2**, where it is assumed that there are four outputs from 0 to 3. Taking the operation of sub-BRB1 as an example, sub-BRB1 receives samples within the 0-1 range from the main classifier. At this point, the output of the main classifier is a fuzzy assessment of the overall system fault status. The role of sub-BRB1 is to conduct a more in-depth analysis based on this assessment, in order to discriminate between fault 1 and fault 2 more precisely.

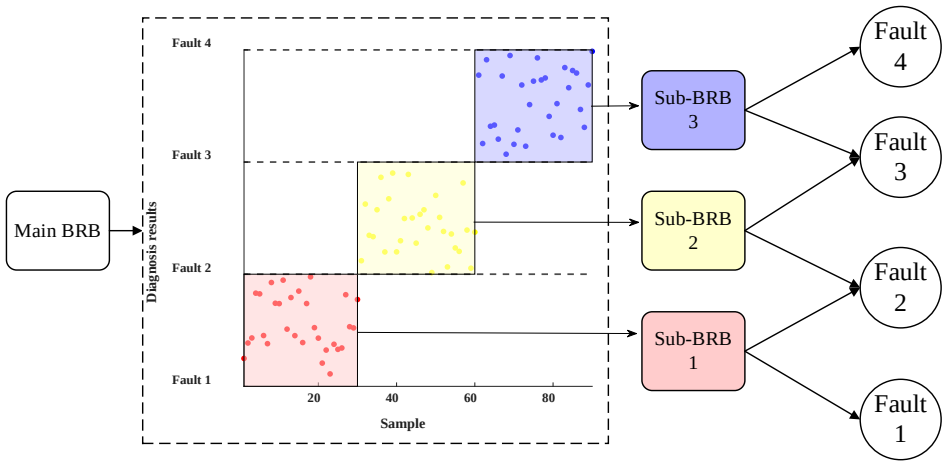

Fig. 2. The construction of the HFS-IBRB model

This hierarchical structure design fully utilizes the benefits of the BRB in handling multiclass and binary classification problems [16]. The main BRB provides global fault category diagnostic information, while the sub-BRBs offer more detailed classification results for each fault category. Such a hierarchical analysis approach helps enhance the awareness of various fault scenarios, thereby improving the accuracy and reliability. Additionally, by combining the main BRB and sub-BRBs, the entire HFS-IBRB model can address class imbalance while maintaining its versatility and flexibility. Each sub-BRB focuses on classifying two specific faults, which weakens the negative impact caused by class imbalance. Meanwhile, it disperses a large number of fault features, reducing the risk of rule explosion.

The number of rules in the BRB is calculated using $R_{num}=\prod_{i=1}^{M}T_i$, where T_i represents the number of reference values of X_i. When too many antecedent attributes are used, the number of rules in a single-layer BRB model rapidly increases, leading to a sharp rise in computational complexity, as shown in **Fig. 3**. Suppose there are a total of 5 features and each feature is assigned 4 reference values. If these features are input into the traditional BRB at once, the number of rules is $\prod_{i=1}^{5}4$. This situation can result in significant time and computational resource waste during the training and inference processes of the model, thereby reducing the efficiency and practicality. Moreover, the rule explosion can also lead to a decrease in the

interpretability, as the complex combinations of rules become difficult to intuitively understand and explain. However, if HFS-IBRB model is used to deal with fault diagnosis under five features, the number of rules can be reduced to $B\prod_{i=1}^{2}4$, where B represents the number of all models in HFS-IBRB, equal to the number of fault types. Through this method, the reduction in the number of rules effectively resolves the issue of rule explosion.

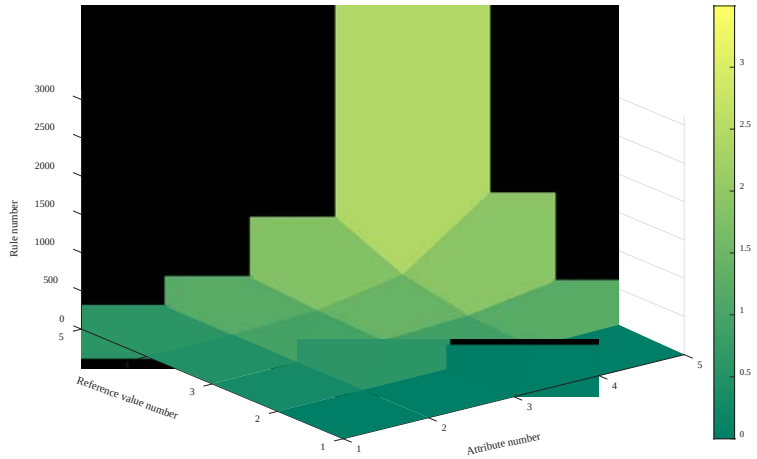

Fig. 3. Rule explosion

3.2.2 XGBoost Feature selection

XGBoost is an efficient and flexible machine learning algorithm proposed by Chen in 2016 [18]. XGBoost can automatically calculate the importance scores of features. Through feature importance evaluation, it can identify which features are most critical for predicting the target, which helps in making better feature selection and data understanding decisions.

Assume that the dataset that requires feature selection contains features and a total of samples. The description of the output of XGBoost is as follows:

$$\hat{y}_i = \sum_{t=1}^{T} f_t(x_i),\ f_t \in F \tag{6}$$

where $\hat{y}_i$ represents the predicted value, f_t represents the structure of a decision tree, and F represents the tree space.

Equation (7) depicts the objective function of XGBoost, while (8) and (9) delineate the loss function and gain function, respectively.

$$obj = \sum_{i=1}^{n} L(y_i, \hat{y}) + \sum_{i=1}^{t} \Omega(f_i) \tag{7}$$

$$L^{(t)} = \sum_{i=1}^{k} \left[l\left(y_i, y_i^{t-1} + g_i f_t(x_i) + \frac{1}{2} h_i f_t^{\,2}(x) \right) \right] + \Omega(f_t) \tag{8}$$

$$gain = \frac{1}{2}\left[\frac{(\sum_{\in I_L} g_i)^2}{\sum_{\in I_L} h_i + \lambda} + \frac{(\sum_{\in I_R} g_i)^2}{\sum_{\in I_R} h_i + \lambda} - \frac{(\sum_{\in I} g_i)^2}{\sum_{\in I} h_i + \lambda} \right] - \gamma \tag{9}$$

Equation (8) is the second-order Taylor series of L at the t-th iteration, $\Omega(f_i)$ represents a regularized term, and g_i and h_i represent the first- and second-order gradients, respectively. $I = I_L \cup I_R$, I_L and I_R represent the samples of the left and right nodes after segmentation, respectively. γ and λ are the penalty parameters.

A greater feature importance score signifies an elevated significance for the associated feature. The two most prominent features, as determined by their feature importance scores, are employed for the construction of each BRB.

3.3 The interpretability of the HFS-IBRB model

The main objective of optimization algorithms is typically to find an optimal or approximately optimal solution within given problem constraints. The stochastic nature of optimization algorithms allows them to explore the solution space more comprehensively, enhancing their ability to perform global searches, particularly when dealing with complex problems to find global optimal solutions. Nevertheless, this inherent characteristic may result in the optimization model losing its physical meaning and interpretability when optimizing the parameters and structure of BRB [19]. The lack of interpretability makes it challenging to understand the internal workings of the model, hindering in-depth analysis and correction of performance issues. This can also lead to mistrust from users and stakeholders. Users typically prefer models that offer transparency and strong interpretability, especially in critical scenarios where understanding and accepting model decisions are essential. Therefore, interpretability is crucial for practical applications in fault diagnosis [20].

Cao et al. proposed a set of general interpretability criteria to guarantee that the model can provide predictions or decisions while being able to explain its decision-making process and outcomes in a clear, intuitive, and understandable manner [21]. Therefore, in model design, it is imperative to ensure that the model complies with the interpretability criteria proposed by Cao et al.

A reasonable belief distribution should have a monotonic or convex shape [22], as shown in the first three graphs in **Fig. 4**. However, the belief distribution displayed in the fourth graph in **Fig. 4**, which might occur when too much emphasis is placed on accuracy, is unreasonable. Therefore, during the optimization process, it is crucial to ensure that the generated belief rules better reflect the characteristics of the actual system in terms of their shape.

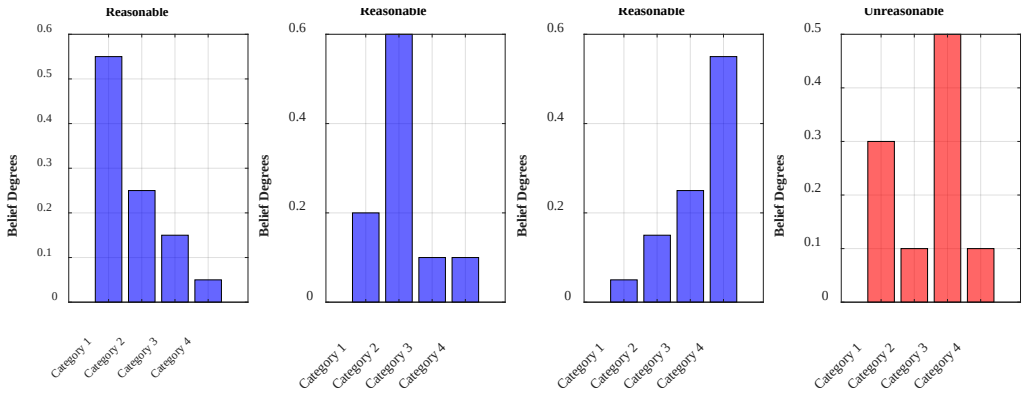

Fig. 4. Comparison of two belief distributions

To ensure that the belief distribution aligns with the mechanism and reality, it is necessary to impose certain constraints [23]. These constraints are designed to regulate the shape and behavior of the belief distribution, making it more interpretable and meaningful in practical applications. Additionally, they help avoid generating unrealistic or unreasonable belief distributions that may arise when overly focusing on optimizing model accuracy at the expense of interpretability. The interpretability constraint designed in this paper is as follows:

$$\begin{aligned} &\beta_k \sim C_k (k = 1, ..., L) \\ &C_k \in \{ \{ \beta_1 \le \beta_2 \le \cdots \le \beta_N \} \\ &\qquad or \{ \beta_1 \ge \beta_2 \ge \cdots \ge \beta_N \} \\ &\qquad or \{ \beta_1 \le \cdots \le \max(\beta_1, \beta_2, \cdots, \beta_N) \ge \cdots \ge \beta_N \} \} \end{aligned} \tag{10}$$

3.4 The inference of the HFS-IBRB model

Due to the complexity of bearings, there is inherent fuzziness and uncertainty in diagnostic processes, necessitating the transformation of indicator data into a unified form that quantitatively describes uncertainty. In light of this, the paper employs the analytical evidential reasoning (ER) algorithm to reason, which effectively integrates multi-index information to assess complex systems, providing support for the quantitative and qualitative expression of information under uncertain conditions. The inference of the HFS-IBRB is summarized as follows:

Step 1: The matching degree a_i^k.

$$a_i^k = \begin{cases} \dfrac{A_i^{l+1} - x_i}{A_i^{l+1} - A_i^l} & k = l, A_i^l \le x_i \le A_i^{l+1} \\ \dfrac{x_i - A_i^{l+1}}{A_i^{l+1} - A_i^l} & k = l+1 \\ 0 & else \end{cases} \tag{11}$$

where A_i^k represents the k-th reference value of the i-th input.

Step 2: The activation weight w_k.

$$w_k = \frac{\theta_k \prod_{i=1}^{N} (a_i^k)^{\delta_i}}{\sum_{i=1}^{K} \theta_l \prod_{i=1}^{N} (a_i^l)^{\delta_i}} \tag{12}$$

Step 3: The belief degree β_n and the utility value μ corresponding to the n-th diagnosis result.

$$\beta_n = \frac{\mu \times [\prod_{i=1}^{L}(w_l \beta_{n,l} + 1 - w_l \sum_{i=1}^{N} \beta_{i,l}) - \prod_{i=1}^{L}(1 - w_l \sum_{i=1}^{N} \beta_{i,l})]}{1 - \mu \times [\prod_{i=1}^{L}(1 - w_l)]} \tag{13}$$

$$\mu = \frac{1}{\sum_{n=1}^{N}\prod_{l=1}^{L}(w_l \beta_{n,l} + 1 - w_l \sum_{i=1}^{N} \beta_{i,l}) - (N-1)\prod_{l=1}^{L}(1 - w_l \sum_{i=1}^{N} \beta_{i,l})} \tag{14}$$

Step 4: The output of the process.

$$y_t = \sum_{n=1}^{N} \mu_n \beta_n \tag{15}$$

3.5 Parameter optimization

In this paper, an optimization algorithm with interpretability constraints is designed based on the projection covariance matrix adaptive evolutionary strategy (P-CMA-ES) optimization algorithm. Before using the optimization algorithm, it is necessary to determine an objective function. The design of the function in this paper is as follows:

$$\begin{aligned} &\min \ MSE(\varphi) \\ &s.t. \sum_{n=1}^{N} \beta_{k,n} = 1 \\ &0 \le \theta_k \le 1 \qquad k = 1, \ldots, R \\ &0 \le \beta_{k,n} \le 1 \qquad n = 1, \ldots, 2^M \\ &0 \le \delta_i \le 1 \qquad i = 1, \ldots, N \\ &\beta_k \sim G_k \end{aligned} \tag{16}$$

where MSE stands for the mean square error, and its calculation is given by (17).

$$MSE = \frac{1}{t}\sum_{i=1}^{t}(y_i - \dot{y}_i)^2 \tag{17}$$

where t is the total number of training samples, y_i is the real value, and $\dot{y}_i$ is the fault diagnosis result of the HFS-IBRB model.

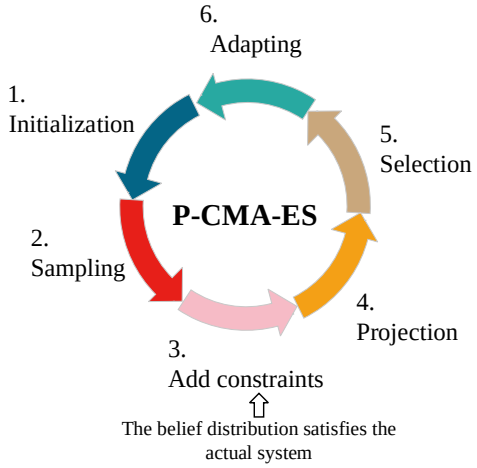

Fig. 5. Optimization process

The optimization algorithm with interpretability constraints used in the HFS-IBRB model is illustrated in **Fig. 5**, and its steps are as follows:

Step 1: Initialization.

Let the initial parameter be denoted as $\Phi^0 = \varphi^0$, where φ^0 represents the initial parameter vector that requires optimization. The parameter set φ is defined as follows:

$$\varphi = [\theta_1, ..., \theta_R, \beta_{1,1}, ..., \beta_{R,N}, \delta_1, ..., \delta_N] \tag{18}$$

Step 2: Sampling.

To generate the population, the following operations need to be performed:

$$\varphi_v^{t+1} = \Phi^t + \gamma^t \varepsilon(0, M^t), v = 1, ..., h \tag{19}$$

where φ_v^{t+1} is the v-th solution vector generated for the (t+1)-th time, Φ represents the mean of the population, γ is the step size, ε is the normal distribution, and M^t is the covariance matrix of the t-th generation population.

Step 3: Add constraints.

Incorporating interpretable constraints makes the HFS-IBRB model more interpretable, guaranteeing the coherence of the belief distribution and parameter range. This step is performed as follows:

$$\beta_k^{t+1} \sim C_8, k = 1, 2, ..., d \tag{20}$$

where β_k^{t+1} is the newly generated belief distribution, which fulfills the eighth interpretability criterion C_8.

Step 4: Projection.

To make the solution vector meet the constraints of optimization, this operation includes the following:

$$\begin{aligned}&\varphi_i^{t+1}(1+j\times(\tau-1):j\times\tau)\\&=\varphi_i^{t+1}(1+j\times(\tau-1):j\times\tau)-L^T\times(L\times L^T)^{-1}\\&\times\varphi_i^{t+1}(1+j\times(\tau-1):j\times\tau)\times L\end{aligned}\tag{21}$$

where $j=(1...J)$ represents the number of variables with constraints, τ denotes the number of constraints, and $L=[1,1,...,1]_{1\times R}$ is the parameter vector.

Step 5: Selection.

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

$$\Phi^{t+1}=\sum_{i=1}^{S}q_i\Phi_i^{t+i},\sum_{i=1}^{S}q_i=1\tag{22}$$

where q_i is the weight coefficient of the i-th solution in the optimal subgroup, $1\le i\le S$.

Step 6: Adapting.

Execute the adapting operation to update the covariance matrix.

$$\begin{aligned}&M^{t+1}=(1-e_1-e_S)T^t+e_1s_c^{t+1}(s_c^{t+1})^T+\\&e_S\sum_{i=1}^{S}t_i(\frac{\varphi_i^{t+1}-\Phi^t}{\gamma^t})\times(\frac{\varphi_i^{t+1}-\Phi^t}{\gamma^t})^T\end{aligned}\tag{23}$$

$$s_c^{t+1}=(1-e_c)s_c^t+\sqrt{e_c(2-e_c)(\sum_{i=1}^{S}t_i^2)^{-1}}\times\frac{\Phi^t\Phi^{t+1}}{\tau^t}\tag{24}$$

$$\tau^{t+1}=\tau^t\exp(\frac{e_\varepsilon}{O_\varepsilon}(\frac{\|s_c^{g+1}\|}{\|H(0,I_m)\|}-1))\tag{25}$$

$$\begin{aligned}&s_\tau^{t+1}=(1-e_\varepsilon)s_\varepsilon^t+\sqrt{e_c(2-e_c)(\sum_{i=1}^{S}t_i^2)^{-1}}\times\\&M^{t-\frac{1}{2}}\times\frac{\Phi^{t+1}-\Phi^t}{\tau^t}\end{aligned}\tag{26}$$

where e_1 , e_S , e_c , e_η represent learning rates, s_τ^t represents the t-th evolutionary step, and $s_\tau^t=0$. In addition, I_m represents the identity matrix, o_η represents the damping coefficient, and $H(0,M^t)$ represents the mathematical expectation.

The entire process will achieve the best interpretability and diagnosis performance in the specific context of the current problem. By gradually optimizing the parameters and covariance matrix, the resulting HFS-IBRB model will possess improved generalization capability, enabling more accurate handling of new input data.

4. Case study

In Section 4.1, relevant information about the datasets used to validate the effectiveness of the model is presented. In Section 4.2, the modeling process is carried out for the given dataset. In Sections 4.3, the model interpretability is verified. In Section 4.4, the performance under class imbalance is verified. In Section 4.5, the advantages of the model are verified by contrast experiments.

4.1 Dataset Background

In this section, the bearing dataset from Southeast University is selected for experimental verification. These data were obtained from the Drivetrain Dynamic Simulator (DDS). The dataset includes five types of fault results: ball fault (B), inner ring fault (I), outer ring fault (O), combination fault on both inner ring and outer ring (C), and health state (H) [24]. In this paper, the dataset is used under the condition of a speed load set to 20 Hz-0 V. In this particular condition, there are eight channels of data, as illustrated in **Table 1**.

Table 1. Description of eight channels

Channel	Description
1	Motor vibration (M_v)
2	Planetary gearbox vibration in the x direction (P_x)
3	Planetary gearbox vibration in the y direction (P_y)
4	Planetary gearbox vibration in the z direction (P_z)
5	Motor torque (M_t)
6	Vibration parallel gearbox vibration in the x direction (V_x)
7	Vibration parallel gearbox vibration in the y direction (V_y)
8	Vibration parallel gearbox vibration in the z direction (V_z)

4.2 Modeling process

The HFS-IBRB model adopts a structure comprising a main BRB and several sub-BRBs. In the main BRB, all faults are diagnosed by classifying the input data into multiple fault types. On the other hand, each individual sub-BRB focuses only on diagnosing two specific fault features. The data contain five fault categories; therefore, the main BRB performs a five-class classification task, while the sub-BRBs sequentially perform four binary classification tasks.

Step 1: Set the reference values for the output of the main BRB.

According to expert knowledge, the five fault categories of the bearings in the dataset {H, C, O, I, B} are assigned reference values {0, 1, 2, 3, 4}.

Step 2: Determining the objectives of the sub-BRBs.

Since each sub-BRB is responsible for handling the classification task of two specific fault categories, four sub-BRBs are required when dealing with five fault categories. Each sub-BRB is designed to model different fault combinations, ensuring that the model can effectively classify all five fault categories.

Sub-BRB1 performs the diagnostic task for faults H and C, sub-BRB2 handles the diagnostic task for faults C and O, and so on. The last sub-BRB, which is the fourth sub-BRB, is responsible for the diagnostic task of faults I and B.

Step 3: Train data set segmentation.

By dividing the training data set into different sub-data sets, each sub-BRB model can focus on specific binary classification tasks, enabling better learning and adaptation to their respective fault features. It is essential to emphasize that the main BRB is trained using the entire data set, ensuring that it can simultaneously handle classification tasks for multiple fault categories, thereby achieving comprehensive fault diagnosis.

Step 4: Feature selection

Feature selection is performed to choose two features for each BRB. **Table 2** presents the chosen pair of features for each BRB, along with their corresponding BRB.

Table 2. Selected features

Model	The 1st feature	The 2nd feature
Main BRB	M_v	M_t
Sub-BRB1	M_v	P_x
Sub-BRB2	M_v	P_x
Sub-BRB3	M_v	M_t
Sub-BRB4	M_t	M_v

Step 5: Model construction and optimization

(1) Main BRB

In the entire dataset, preliminary diagnosis is conducted for the five categories of bearing faults. Based on expert knowledge, the two features selected through feature selection are assigned typical semantic values, including VL (very large), L (large), M (middle), and S (small), as shown in **Table 3**.

Table 3. The reference values

Attribute	VL	L	M	S
M_v	-0.1	-0.13	-2.6	-2.7
M_t	0.023	0.0025	-0.003	-0.029

After the preliminary diagnosis by the main BRB, the diagnostic results for the five categories of bearing faults are shown in **Fig. 6 (1)**. The confusion matrix in **Fig. 6 (1)** provides a detailed statistical summary of the classification results of the main BRB on different categories. The accuracy of 81.33% reflects that the model performs well in classifying the majority of samples but also indicates errors in a portion of samples. Analyzing the model's diagnostic performance in each class through the confusion matrix reveals that the classification ability among similar fault categories still needs improvement. Therefore, further optimization and adjustments to the model can help enhance the accuracy.

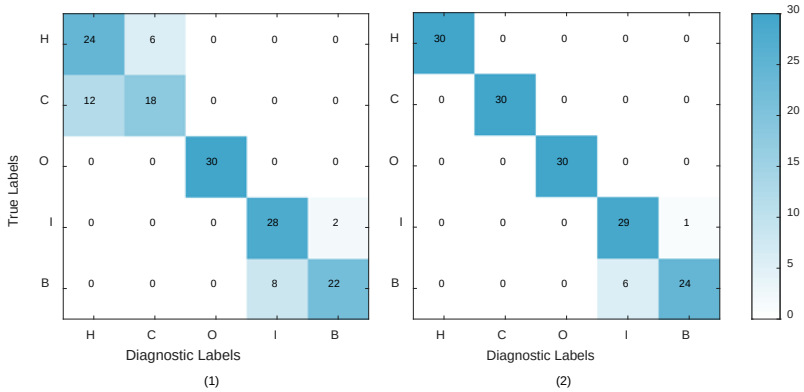

Fig. 6. Results of the main BRB and the overall HFS-IBRB model

(2) Sub-BRBs

To further refine the fault diagnosis process, the HFS-IBRB model implements the diagnosis of sub-BRBs based on the main BRB. The main BRB is responsible for the preliminary classification of all faults, while each sub-BRB focuses on specific fault combinations, diagnosing two specific fault categories. In addition, in fault diagnosis problems, there exists a transitional phase between different fault categories, where neighboring categories exhibit high attribute similarity. This can give rise to the generation of local ignorance information, potentially leading to the model inaccurately reflecting objective facts. Hence, within each sub-BRB, to effectively represent ignorance information, an intermediate state is introduced, and confidence degrees are allocated to three states. This approach better captures local ignorance and enhances the representation.

Fig. 6 (2) below integrates the fault diagnosis results from the four sub-BRBs and displays the model's diagnostic performance in the form of a confusion matrix. According to the illustration in **Fig. 6 (2)**, it can be inferred that through the refinement by the sub-BRBs, the diagnostic accuracy of the model has increased from 81.33% to 95.33%.

4.3 The interpretability of the HFS-IBRB model

To verify the effectiveness of the interpretability constraints, the belief degree of expert knowledge, the optimized HFS-IBRB model and the hierarchical BRB model based on XGBoost feature selection (HFS-BRB) are compared. Take the main BRB as an example, and its comparison effect is shown in **Fig. 7**. After the general BRB model was optimized, its parameter structure was disrupted, leading to significant discrepancies between the belief degree and expert knowledge, particularly evident in cases such as Rule 5 in **Fig. 7**. However, the proposed HFS-IBRB model exhibited a high degree of similarity between its belief degree and expert knowledge, retaining more of the expert knowledge. This suggests that the model possesses strong interpretability, and the implementation of the interpretability constraints holds practical significance for the model.

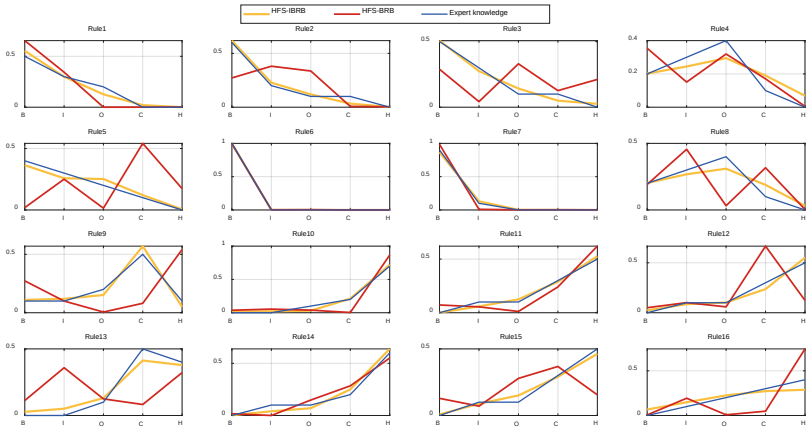

Fig. 7. The comparison of belief distribution

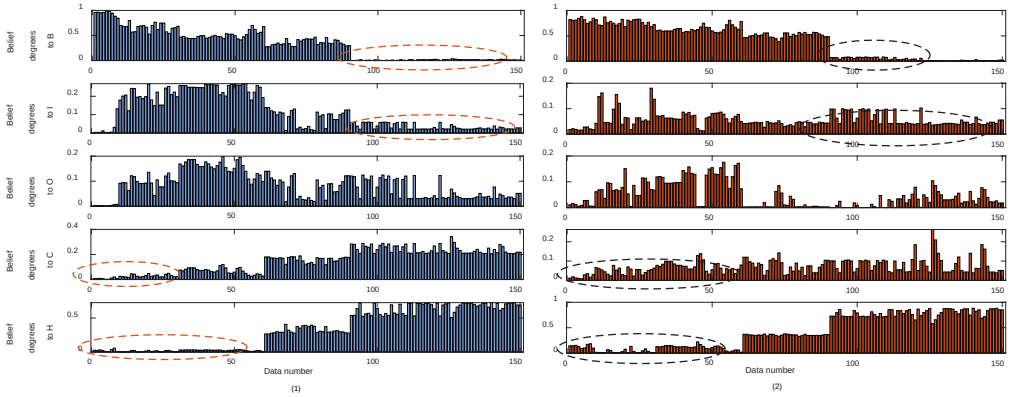

Fig. 8. The distributional results generated by HFS-IBRB and HFS-BRB

Fig. 8(1) and **Fig. 8(2)** compare the belief distribution results of the main BRBs of the two hierarchical models, including HFS-IBRB and HFS-BRB. The HFS-IBRB model offers a clear semantic representation of the diagnostic results. By observing the red dashed boxes in both figures, it is evident that the degree of support for the results within the box of HFS-IBRB is smaller than that within the box of **Fig. 8(2)**. Therefore, HFS-IBRB possesses a stronger capability to handle uncertainty, enabling more accurate fault diagnosis.

4.4 Analysis of unbalanced classification ability

Under the background of bearing fault diagnosis, the ability to address imbalanced classification issues is of paramount importance. In real-world scenarios, the challenge of imbalanced diagnostic categories frequently arises, wherein certain fault categories may exhibit a scarcity of instances compared to others. This imbalance can result in suboptimal diagnostic performance for minority classes. In this section, the dataset proportions are adjusted to 1:2:3:4:5, and on this foundation, the diagnostic performance of the main BRB and the overall HFS-IBRB is shown in **Fig. 9 (1)** and **Fig. 9 (2)**.

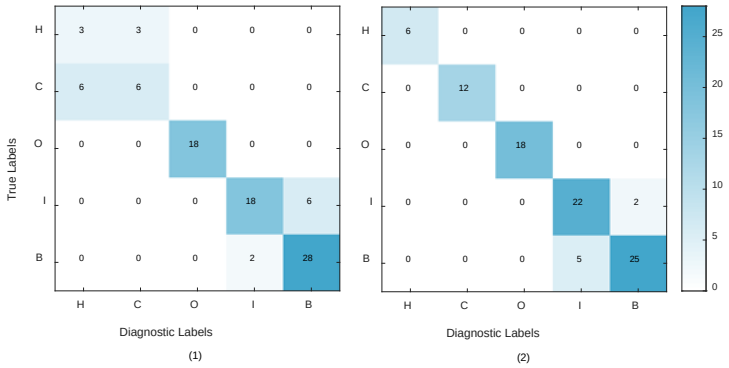

Fig. 9. Results of the main BRB and the overall HFS-IBRB model under class imbalance

In the context of class imbalance, the proposed bearing fault diagnosis model based on HFS-IBRB achieves a diagnostic accuracy of 92.22%. This clearly highlights the distinct advantage of the HFS-IBRB model in addressing class imbalance issues. Firstly, BRB, serving as the foundation of the HFS-IBRB model, capitalizes on its robust non-linear modeling capability and effective integration of expert knowledge, which contributes to a better handling of uneven data distribution among different classes. By infusing expert knowledge into the model, it becomes capable of more accurately capturing the correlations and characteristics among various fault categories, thereby enhancing the diagnostic ability for minority classes. Secondly, the HFS-IBRB model proposed in this paper employs a two-level framework. In the first level, the main BRB approximates the classification of confusing categories, followed by transmission to multiple secondary BRBs in the second level for binary classification. This sequential process ultimately improved the accuracy from 81.11% of the main BRB to 92.22%.

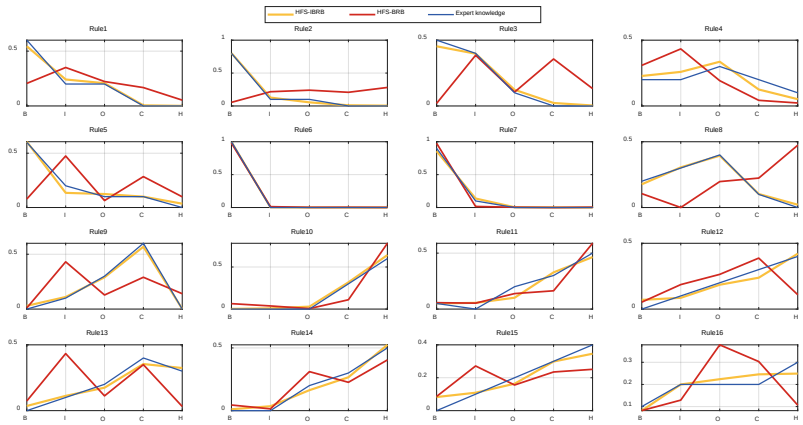

Fig. 10. The comparison of main BRB

The HFS-IBRB model not only enhances the accuracy of the diagnostic process but also provides a clear and transparent diagnostic procedure. Take the main BRB as an example, and its comparison effect of three different belief distribution scenarios is shown in **Fig. 10**. The belief distribution obtained by the HFS-IBRB model in **Fig. 10** exhibits a high degree of similarity to expert knowledge. This suggests that the rules generated by the HFS-IBRB model are consistent with the practical fault diagnostic system. However, certain rules in the diagram, such as Rule 7, although aligned with expert knowledge, show a slightly lower degree of similarity compared to the HFS-IBRB model.

4.5 Contrast experiment

To objectively assess the effectiveness of the model, this section selected Backpropagation Neural Network (BPNN), radial basis function (RBF), and Decision Tree for comparative experiments under the condition of unbalanced classes. The experiments were repeated for 20 rounds, as shown in **Fig. 11**. In **Table 4**, Part 1 compares the accuracy of various BRB models used for bearing fault diagnosis, while Part 2 shows the accuracy of other models. Although the accuracy of HFS-IBRB and HFS-BRB model is similar, HFS-IBRB model is more interpretable, so it has more application value.

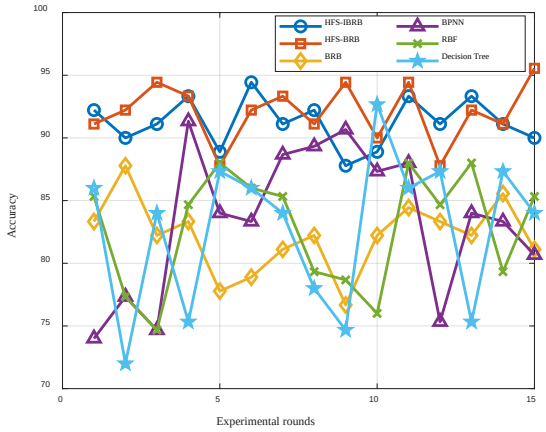

Fig. 11. The comparison of performance among different methods

Table 4. Comparison of different models

Part	Method	Max accuracy (%)	Min accuracy (%)	Average accuracy (%)	The standard deviation of accuracy (%)
Part 1	HFS-IBRB	94.44	87.78	91.26	1.85
	HFS-BRB	95.55	87.78	92.07	2.25
	BRB	87.78	76.67	82.15	2.77
Part 2	BPNN	91.33	74.00	83.47	5.72
	RBF	88.00	74.67	82.71	4.48
	Decision Tree	92.67	72.00	82.67	5.84

The hierarchical BRB model with interpretability developed based on BRB demonstrates significant advantages in dealing with the bearing fault diagnosis in the case study:

(1) Compared to the traditional BRB, the HFS-IBRB shows greater accuracy and scalability. The HFS-IBRB establishes a hierarchical structure that enables it to flexibly address class imbalance issues. In practical case studies, where certain fault categories have relatively few samples, the traditional BRB may perform poorly in diagnosing these categories. In contrast, the HFS-IBRB optimizes each binary classification task, effectively capturing distinct features and associations between different categories, thus mitigating the impact of class imbalance.

(2) The HFS-IBRB offers excellent interpretability, with a transparent inference process and traceable outcomes. Unlike data-driven black-box models, the HFS-IBRB presents the decision-making process and reasoning logic clearly, allowing users to understand how diagnostic results are obtained. This high level of interpretability not only enhances user trust in the model but also provides valuable guidance for further model optimization and diagnostic

process improvement.

(3) The application of hierarchical structure in complex system modeling extends beyond addressing the issue of class imbalance. It also effectively tackles the problem of rule explosion. The hierarchical design prevents the exponential combination of rules as the rule count increases, thereby reducing the complexity of computation and inference. The rules within the hierarchical structure are more targeted and aligned with the characteristics of the actual problem, enabling more precise inference and decision-making.

In conclusion, the HFS-IBRB developed based on BRB demonstrates clear advantages in addressing class imbalance issues in the case study. Leveraging its hierarchical model structure and high interpretability, the HFS-IBRB model not only enhances diagnostic accuracy but also provides a reliable and comprehensible solution to complex bearing fault diagnosis problems.

5. Conclusion

This paper proposes an HFS-IBRB model for bearing fault diagnosis. Firstly, this paper adopts a hierarchical structure to address the class imbalance. Secondly, the analytical ER algorithm is utilized for inference. Lastly, a P-CMA-ES-based optimization algorithm with interpretability constraints is proposed through the design of interpretability constraints.

The HFS-IBRB model consists of a main BRB and multiple sub-BRBs. The main BRB is responsible for the preliminary classification of all faults, while each sub-BRB focuses on specific fault combinations, that is, diagnosing two specific fault categories. With this design, the model can more effectively handle multi-class fault diagnosis problems with unbalanced classes. The emphasis on interpretability in this paper also enhances the trustworthiness. The achievements of this study are not only in proposing a bearing fault diagnosis model, which demonstrates a notable improvement in diagnostic accuracy for class imbalance issues, but also in showcasing the effectiveness of BRB as a valuable tool to mitigate the challenges posed by class imbalance in the field of fault diagnosis. This provides robust support for practical applications in fault diagnosis and offers valuable insights for addressing class imbalance challenges.

During the research process of the HFS-IBRB model, some issues have been identified for future investigation. Firstly, improving the accuracy of the main BRB is crucial for enhancing the overall system performance. The accurate classification by the main BRB directly influences the subsequent judgments of the sub-BRBs. Therefore, investigating how to ensure high accuracy at the main BRB stage is a primary focus of the upcoming research. Secondly, complex systems like bearings are often subject to environmental disturbance during operation, such as vibration and temperature changes, which can impact monitoring data. Exploring ways to enhance anti-interference and quantify uncertainty is also the focus of future research.

Acknowledgement

This work was supported in part by the Postdoctoral Science Foundation of China under Grant No. 2020M683736, in part by the Teaching reform project of higher education in Heilongjiang Province under Grant No. SJGY20210456, in part by the Natural Science Foundation of Heilongjiang Province of China under Grant No. LH2021F038, in part by the Social Science Foundation of Heilongjiang Province of China under Grant No. 21GLC189, in part by the Foreign Expert Projects in Heilongjiang under Grant No. GZ20220131, in part by the Graduate innovation project of Harbin Normal University under Grant No.HSDSSCX2023-3.

References

[1] Liu Z H, Chen L, Wei H L, et al., "A Tensor-based domain alignment method for intelligent fault diagnosis of rolling bearing in rotating machinery," *RELIAB ENG SYST SAFE*, vol. 230, p. 108968, 2023. Article (CrossRef Link)

[2] Chen X, Yang R, Xue Y, et al., "Deep transfer learning for bearing fault diagnosis: A systematic review since 2016," *IEEE T INSTRUM MEAS*, vol. 72, pp. 1-21, 2023. Article (CrossRef Link)

[3] Tao H, Qiu J, Chen Y, et al., "Unsupervised cross-domain rolling bearing fault diagnosis based on time-frequency information fusion," *J FRANKLIN I*, vol. 360, no. 2, pp. 1454-1477, 2023. Article (CrossRef Link)

[4] Rzadkowski R, Rokicki E, Piechowski L, et al., "Analysis of middle bearing failure in rotor jet engine using tip-timing and tip-clearance techniques," *MECH SYST SIGNAL PR*, vol. 76-77, pp. 213-227, 2016. Article (CrossRef Link)

[5] Wang Xing, Zhang Han, Zhu Jiazheng, et al., "Fault diagnosis method of aviation high-speed bearing driven by multi-heads," *Journal of Vibration and Shock*, vol. 42, no. 4, pp. 295-305, 2023. Article (CrossRef Link)

[6] Tai C Y, Altintas Y., "A hybrid physics and data-driven model for spindle fault detection," *CIRP ANN*, vol. 72, no. 1, pp. 297-300, 2023. Article (CrossRef Link)

[7] Fu L, Cheng L, Chen W, et al., "Bearing Fault Diagnosis with Small Sample Based on Data-Driven," in *Proc. of 2023 4th ICTC. IEEE*, pp. 369-373, 2023. Article (CrossRef Link)

[8] Chen Y W, Yang J B, Xu D L, et al., "On the inference and approximation properties of belief rule based systems," *INFORM SCIENCES*, vol. 234, pp. 121-135, 2013. Article (CrossRef Link)

[9] He W, Cheng X, Zhao X, et al., "An interval construction belief rule base with interpretability for complex systems," *EXPERT SYST APPL*, vol. 229, p. 120485, 2023. Article (CrossRef Link)

[10] Yang J B, Liu J, Wang J, et al., "Belief rule-base inference methodology using the evidential reasoning approach-RIMER," *IEEE T SYST MAN CY A*, vol. 36, no. 2, pp. 266-285, March. 2006. Article (CrossRef Link)

[11] Manlin C, Zhijie Z, Zhang B, et al., "A novel combination belief rule base model for mechanical equipment fault diagnosis," *CHINESE J AERONAUT*, vol. 35, no. 5, pp. 158-178, 2022. Article (CrossRef Link)

[12] Zhou Z, Cao Y, Hu G, et al., "New health-state assessment model based on belief rule base with interpretability," *SCI CHINA INFORM SCI*, vol. 64, no. 7, p. 172214, 2021. Article (CrossRef Link)

[13] Zhou Z J, Hu G Y, Hu C H, et al., "A survey of belief rule-base expert system," *IEEE T SYST MAN CY*, vol. 51, no. 8, pp. 4944-4958, Aug. 2021. Article (CrossRef Link)

[14] Chen Y W, Yang J B, Pan C C, et al., "Identification of uncertain nonlinear systems: Constructing belief rule-based models," *Knowledge-Based Systems*, vol. 73, pp. 124-133, 2015. Article (CrossRef Link)

[15] Feng Z, Zhou Z J, Hu C, et al., "A new belief rule base model with attribute reliability," *IEEE T FUZZY SYST*, vol. 27, no. 5, pp. 903-916, May. 2019. Article (CrossRef Link)

[16] Hu G, He W, Sun C, et al., "Hierarchical belief rule-based model for imbalanced multi-classification," *EXPERT SYST APPL*, vol. 216, p. 119451, 2023. Article (CrossRef Link)

[17] Cao Y, Zhou Z J, Hu C H, et al., "A new approximate belief rule base expert system for complex system modelling," *DECIS SUPPORT SYST*, vol. 150, p. 113558, 2021, Article (CrossRef Link)

[18] Chen T, Guestrin C., "Xgboost: A scalable tree boosting system," in *Proc. of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp.785-794, 2016. Article (CrossRef Link)

[19] Feng Z, Zhou Z, Hu C, et al., "A safety assessment model based on belief rule base with new optimization method," *RELIAB ENG SYST SAFE*, vol. 203, p. 107055, 2020. Article (CrossRef Link)

[20] Zhichao M, Zhijie Z, You C A O, et al., "A new interpretable fault diagnosis method based on belief rule base and probability table," *CHINESE J AERONAUT*, vol. 36, no. 3, pp. 184-201, 2023. Article (CrossRef Link)

[21] Cao Y, Zhou Z, Hu C, et al., "On the interpretability of belief rule-based expert systems," *IEEE T FUZZY SYST*, vol. 29, no. 11, pp. 3489-3503, Nov. 2021. Article (CrossRef Link)
[22] Han P, He W, Cao Y, et al., "Lithium-ion battery health assessment method based on belief rule base with interpretability," *APPL SOFT COMPUT*, vol. 138, p. 110160, 2023. Article (CrossRef Link)
[23] Cao Y, Zhou Z, Tang S, et al., "On the Robustness of Belief-Rule-Based Expert Systems," *IEEE T SYST MAN CY*, vol. 53, no. 10, pp. 6043-6055, Oct. 2023. Article (CrossRef Link)
[24] Shao S, McAleer S, Yan R, et al., "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE T IND INFORM*, vol. 15, no. 4, pp. 2446-2455, April. 2019. Article (CrossRef Link)

Boying Zhao received the B.Eng. degree from the 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, fault diagnosis, and interpretable diagnosis systems.

Yuanyuan Qu graduated from Harbin University of Science and Technology with a Master's degree. She is currently working at Heilongjiang Agricultural Engineering Vocational College in Harbin. Her research interests include evidential reasoning and belief rule base.

Mengliang Mu graduated from East China Normal University of Science and Technology with a Master's degree. He is currently working at Heilongjiang Agricultural Engineering Vocational College in Harbin. His research interests include evidential reasoning and deep learning.

Bing Xu received a doctor's degree of Agricultural System Engineering and Management Engineering from Northeast Agricultural University, Harbin, China, in 2018. She is currently 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.

Wei He received a doctor's degree from Harbin University of Science and Technology in 2018. He is currently working at the School of Computer Science and Information Engineering, Harbin Normal University. His research interests include evidential reasoning, deep learning, and belief rule base.