machines

MDPI

Article

Relation Knowledge-Guided Federated Model Compression for Rare-Fault Preservation in Motor Fault Diagnosis

Genbao Zhao and Juan Zhang *

Automation with the College of Information Science and Engineering, Northeastern University, Shenyang 110000, China
* Correspondence: zhangjuan@ise.neu.edu.cn

Abstract

To address global knowledge bias, weak rare-fault recognition, and high edge-deployment costs caused by heterogeneous sample sizes, data quality, fault categories, and monitoring modalities among multiple clients, this paper proposes a rare-fault-preserving federated dynamic model slimming method based on relational knowledge. The core idea is to formulate lightweight federated diagnosis as a joint optimization problem of rare-fault knowledge preservation and redundant knowledge suppression. At each local client, output-discriminative knowledge, class-prototype relations, and input-sensitive relations are extracted to describe diagnostic knowledge from the decision, structure, and weak-response levels. At the federated server, a rare-fault-aware weighting mechanism adjusts the contribution of local knowledge according to sample scarcity, output reliability, and distribution dispersion and then fuses multi-granularity relational knowledge to optimize the global teacher model. A relation-constrained gated slimming strategy is further designed for the student model, enabling the lightweight model to retain critical diagnostic channels while suppressing repetitive and low-contribution information. Experiments on the CWRU bearing dataset and the HUST multimodal motor dataset show that the proposed method achieves higher diagnostic accuracy, rare-fault recall, and deployment efficiency under composite imbalance, cross-condition generalization, and modality-missing deployment scenarios. These results demonstrate the effectiveness of the proposed method for raw-data-free and privacy-aware multi-client motor fault diagnosis.

Keywords: federated learning; model slimming; rare fault; motor fault diagnosis

Check for updates

Academic Editors: Ronghua Zhu, Zepeng Liu, Chaoge Wang and Cailiang Zhang

Received: 14 May 2026
Revised: 8 June 2026
Accepted: 11 June 2026
Published: 15 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Motors are key actuating components in industrial transmission equipment and intelligent manufacturing systems, and their operating conditions directly affect production safety, operational continuity, and maintenance cost [1,2]. During long-term service, motors are subject to load fluctuations, start–stop impacts, installation deviations, component aging, and environmental disturbances, which may lead to various faults, such as bearing damage, rotor unbalance, misalignment, broken rotor bars, winding short circuits, and voltage imbalance [3]. However, motor fault diagnosis in engineering applications is not a simple classification problem but is jointly affected by data decentralization, class scarcity, distribution heterogeneity, privacy constraints, and limited resources for edge deployment.

In multi-client collaborative diagnosis scenarios, differences in motor types, operating conditions, sensor arrangements, and maintenance states lead to significant inconsistencies among local datasets. Some clients may operate mainly under normal or slightly degraded

conditions for long periods, resulting in insufficient fault samples. Incipient and rare faults are usually characterized by weak impacts, strong noise interference, ambiguous class boundaries, and limited sample coverage [4]. Existing studies still have three main limitations in balancing rare-fault preservation and lightweight deployment. First, conventional federated aggregation and federated distillation tend to emphasize global accuracy or knowledge consistency, but rare-fault knowledge from small-sample clients can still be weakened during server-side fusion. Second, most distillation methods rely mainly on output responses or feature representations, which are insufficient to jointly describe soft decision boundaries, inter-class fault structures, and weak-fault sensitivity. Third, existing federated compression or slimming methods mainly reduce parameters or channels, but they do not explicitly prevent the removal of weak yet critical representations related to rare-fault boundaries. As a result, rare-fault features can be easily overwhelmed by majority classes and highly repetitive samples, making it difficult for diagnostic models to preserve sensitivity to minority faults under imbalanced data conditions [5]. Centralized deep learning methods can train diagnostic models by aggregating data from different devices [6], but direct data sharing is often constrained by enterprise privacy policies, equipment security, communication costs, and data management regulations. Federated learning provides a feasible raw-data-free solution for multi-client collaborative modeling by keeping raw monitoring data locally and exchanging model parameters or knowledge descriptors [7]. It should be noted that raw-data-free collaboration does not indicate a formal privacy guarantee because exchanged parameters or descriptors may still contain statistical information about local data. In addition, due to non-IID data distributions, conventional federated aggregation tends to be dominated by clients with larger sample sizes, more complete fault categories, and higher data quality. Consequently, rare-fault knowledge from minority clients may be weakened in the global model [8]. Meanwhile, the global model may also absorb repetitive, low-quality, and weakly task-related knowledge from multiple clients. Such knowledge usually makes a limited contribution to rare-fault discrimination but may increase model complexity and weaken minority-class boundaries. It may arise from repetitive feature responses of large-sample clients, majority-class-dominated decision information, or low-contribution channels that are insensitive to rare and incipient faults. As a result, the global model may contain redundant decision boundaries, leading to higher deployment cost and reduced edge-side applicability.

Pruning, quantization, knowledge distillation, and federated distillation have been widely used to reduce model complexity, storage cost, communication burden, and inference latency in mechanical fault diagnosis [9–12]. However, most existing lightweighting methods are designed for centralized or single-client scenarios, and their objectives mainly focus on parameter compression or teacher-output approximation [13]. Under multi-client imbalanced conditions, minority-fault and incipient-fault responses are usually weak and can be removed during pruning or slimming. Meanwhile, existing federated distillation methods mainly transfer soft labels, feature representations, class prototypes, or generated knowledge to improve model fusion, cross-domain generalization, or communication efficiency, with insufficient attention to rare-fault knowledge preservation during compression [14]. Relying only on output probabilities or conventional feature distillation cannot fully represent inter-class fault structures, input-sensitive weak responses, and complex rare-fault boundaries. Therefore, lightweight federated diagnosis should not only reduce model size but also preserve effective diagnostic knowledge and suppress redundant, low-quality, and majority-class-dominated knowledge.

More specifically, existing related techniques address only part of this requirement. Federated prototype learning aligns feature structures through class prototypes, but it cannot describe soft decision boundaries or input-sensitive weak responses. Federated

knowledge distillation transfers output-level knowledge, while output responses alone are insufficient to preserve structural relations among similar faults. Rare-class reweighting increases the contribution of minority samples during local training, but it cannot evaluate the reliability of rare-fault knowledge during federated fusion. Model slimming reduces structural redundancy, but weak channels associated with rare-fault boundaries may still be removed without explicit knowledge-preservation constraints. Therefore, rare-fault-preserving lightweight federated diagnosis requires a unified framework that jointly considers decision-level discrimination, class-level structural relations, input-level sensitivity responses, rare-fault-aware knowledge fusion, and redundancy-suppressed model slimming.

To address these issues, this paper proposes a rare-fault-preserving federated dynamic model slimming method based on relational knowledge. The core novelty lies in using multi-granularity relational knowledge as a unified constraint for both rare-fault-aware federated fusion and gated student slimming, rather than treating relational distillation, federated weighting, and model slimming as independent modules. Specifically, output-discriminative knowledge, class-prototype relations, and input-sensitive relations are extracted at local clients to describe diagnostic knowledge from the decision, structure, and sensitivity levels. At the federated center, rare-fault-aware weighting is introduced to enhance reliable minority-fault knowledge from clients with scarce samples and heterogeneous data quality. Then, a relation-constrained gated teacher-to-student slimming strategy is constructed to reduce deployment complexity while preserving rare-fault recognition capability. The main contributions of this paper are summarized as follows:

(1) A rare-fault-preserving federated dynamic slimming framework is proposed for multi-client imbalanced motor fault diagnosis. Different from conventional federated distillation or federated compression methods, the proposed framework jointly considers federated knowledge fusion, rare-fault boundary preservation, and redundant channel suppression.

(2) A multi-granularity relational knowledge representation mechanism is developed by integrating output-discriminative knowledge, class-prototype relations, and input-sensitive relations. This design preserves soft decision boundaries, fault-class structural relationships, and weak-fault response sensitivity simultaneously.

(3) A rare-fault-aware knowledge weighting strategy and a relation-constrained gated slimming strategy are designed. The former reduces the dominance of majority clients during federated fusion, while the latter prevents the lightweight student model from removing key channels associated with rare-fault recognition.

2. Related Works

2.1. Lightweight Fault Diagnosis Methods

In mechanical fault diagnosis, lightweight modeling has been widely studied to reduce storage cost, computational burden, and inference latency. Existing methods mainly include parameter pruning, weight quantization, low-rank decomposition, lightweight network design, and knowledge distillation [15,16]. Some studies improve deployment efficiency by using lightweight convolutional networks, depth-wise separable convolutions, compact residual modules, or embedded diagnostic networks. For example, Gong et al. [17] proposed a lightweight fault diagnosis method for embedded systems, and Sun et al. [18] developed a multi-level compression strategy for bearing fault diagnosis. Other studies combine compression with knowledge transfer. Zhang et al. [19] introduced knowledge distillation into fault diagnosis model compression, Ji et al. [20] integrated distillation with parameter quantization, and Liao et al. [21] improved deployment applicability by combining decoupled knowledge distillation with hardware acceleration. These studies show

that lightweight design and distillation can reduce model complexity while maintaining diagnostic performance.

Nevertheless, most existing lightweight fault diagnosis methods are developed under centralized or single-client settings with relatively complete class distributions. Their objectives mainly focus on parameter reduction, inference acceleration, or teacher-output approximation, with insufficient attention to rare-fault preservation during compression. In multi-client motor fault diagnosis, data heterogeneity, inconsistent sample quality, and incomplete fault-category coverage make this problem more severe. When rare faults exist in only a few clients or low-SNR samples, pruning and quantization may remove parameters related to weak fault boundaries, while output distillation alone cannot preserve fault-class structures and weak-response features. Therefore, lightweight diagnosis under multi-client imbalanced conditions should not only reduce model size but also preserve effective fault knowledge and suppress redundant information during compression.

2.2. Knowledge Distillation-Based Fault Diagnosis Methods

Knowledge distillation transfers diagnostic knowledge from a teacher model to a student model and provides an effective compression strategy for lightweight fault diagnosis. Hinton et al. [22] introduced softened output probabilities to help the student model learn inter-class similarity information. Gou et al. [23] further summarized distillation knowledge into response-based, feature-based, and relation-based knowledge, indicating that the effectiveness of distillation depends on both the knowledge type and the transfer strategy. For fault diagnosis, different fault classes often have feature similarities and mechanism-level correlations, so hard-label supervision or output probability fitting alone is insufficient to preserve fine-grained class boundaries. In mechanical fault diagnosis, knowledge distillation has been used for model compression, lightweight deployment, and cross-condition diagnosis. Lu et al. [24] proposed a lightweight distillation transfer learning framework to improve bearing diagnosis under insufficient labeled samples. Other studies introduced intermediate feature hints [25] or relational knowledge distillation [26] to enhance the representation capability of student models. These methods show that distillation can improve lightweight diagnostic models by transferring output, feature, or relation knowledge. However, most existing methods still assume that teacher knowledge is reliable and mainly focus on centralized or single-domain scenarios. Under few-shot, low-SNR, and class-missing conditions, the teacher model may contain both useful discriminative knowledge and noise-induced bias. Directly transferring output responses or intermediate features may fail to preserve class structure relations, inter-sample similarities, and weak rare-fault boundaries.

In addition to knowledge distillation, digital-twin-guided diagnosis has become an important knowledge-driven fault diagnosis paradigm. Digital-twin-guided physical–virtual denoising methods construct bearing dynamic models to generate simulated fault signals and calibrate the virtual model using measured signals, thereby enhancing weak early-fault features under strong background noise [27]. Digital-twin-driven cross-domain adaptation methods further use high-fidelity simulated signals as knowledge sources to reduce the discrepancy between simulated and measured domains [28]. These studies demonstrate the value of generated physical knowledge for improving robustness and interpretability. However, they mainly focus on physical–virtual signal generation, denoising, or simulation-to-real adaptation. In contrast, this study focuses on federated rare-fault knowledge preservation and lightweight model slimming when raw diagnostic data are distributed across multiple clients and cannot be centrally shared.

2.3. Federated Distillation-Based Fault Diagnosis Methods

Federated learning enables multi-client collaborative diagnosis through local training and server-side aggregation, which alleviates data isolation without sharing raw monitoring data. Chen et al. [29] proposed a discrepancy-weighted federated transfer learning method to mitigate cross-client distribution differences in bearing diagnosis. Li et al. [30] investigated a privacy-aware federated transfer diagnosis method with target-adaptive capability. Zhao et al. [31] developed a federated multi-source domain adversarial adaptation framework to reduce distribution shifts among multiple clients. These studies promote the application of federated learning in mechanical fault diagnosis. However, most of them still rely on parameter aggregation or domain adaptation, with limited consideration of model heterogeneity, communication cost, and redundant knowledge accumulation. Federated distillation transfers diagnostic knowledge instead of complete model parameters, providing a more flexible solution for heterogeneous client collaboration and communication reduction. Xue et al. [32] combined federated transfer learning with consensus knowledge distillation, where consensus knowledge was used to guide clients in sharing diagnostic knowledge under raw-data-free collaboration. Sun et al. [13] proposed FedAlign, which uses data-free knowledge distillation and pseudo features to align heterogeneous client models. These studies show that federated distillation can reduce communication burden, support model heterogeneity, and improve cross-client knowledge sharing. However, existing federated distillation-based fault diagnosis methods still have difficulty meeting the requirements of rare-fault preservation and lightweight deployment under multi-client imbalance. Most methods focus on cross-domain generalization, model alignment, or communication efficiency, while paying insufficient attention to how redundant or biased knowledge affects minority-fault recognition. When sample size, sample quality, and fault-category distributions vary across clients, repetitive knowledge from large-sample clients, noise-biased knowledge from low-quality clients, and majority-class-dominated knowledge may weaken rare-fault boundaries. In addition, output probabilities, public-data responses, or pseudo features provide limited characterization of class-relation structures and input-sensitive weak responses. Therefore, multi-client motor fault diagnosis requires a unified design that combines federated distillation, relational knowledge preservation, rare-fault-aware fusion, and dynamic model slimming to support raw-data-free collaboration, rare-fault preservation, redundant knowledge suppression, and edge deployment.

3. Federated Dynamic Slimming Algorithms with Relation Knowledge for Rare-Fault Preservation

The main notation is unified as follows. K denotes the number of clients, C denotes the number of fault classes, D_k denotes the local dataset of client k, and D_a denotes the labeled calibration set at the server. F_k, F_G, and F_S denote the local model, global teacher model, and student model, respectively. Q_k, R_k^p, and R_k^s denote the output knowledge, class-prototype relation, and input-sensitive relation. $\alpha_{k,c}$ denotes the class-level fusion weight, and ξ denotes the channel-retention threshold.

3.1. Overall Framework with Output Discrimination and Relation Knowledge Collaboration

To address the weakening of rare-fault knowledge, the accumulation of redundant knowledge in the global model, and the high deployment complexity at the edge side in multi-cliental motor fault diagnosis, this study constructs a relation knowledge-based federated dynamic slimming framework, as shown in Figure 1. The framework consists of four stages: local knowledge mining, federated knowledge fusion, global teacher model optimization, and gated dynamic slimming of the student model. Different from methods that only upload model parameters or distill output soft labels, the proposed method repre-

sents local diagnostic knowledge using output-discriminative knowledge, class-prototype relations, and input-sensitive relations, and further performs rare-fault-aware fusion at the federated center.

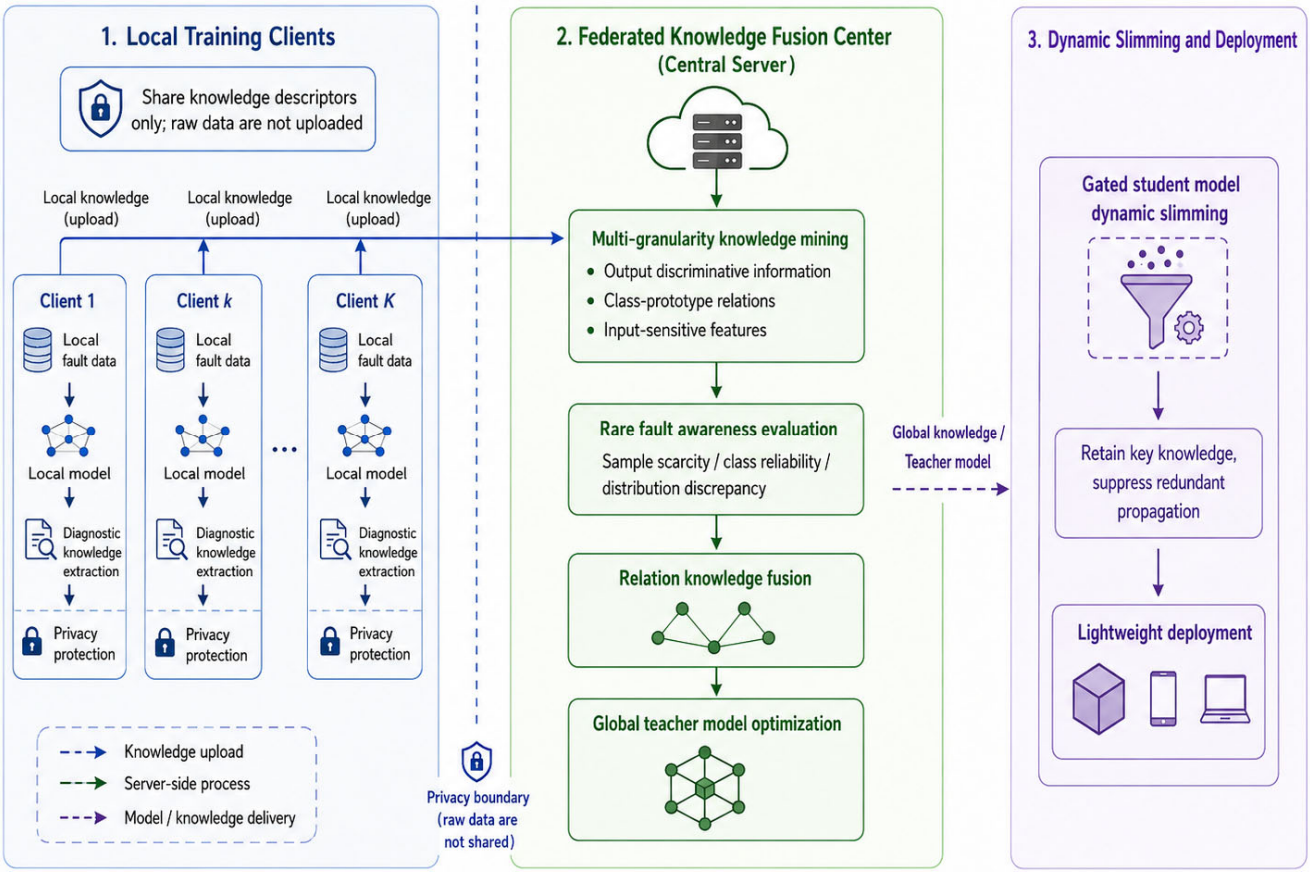

Figure 1. Overall framework of the rare-fault-preserving relation knowledge-based federated dynamic slimming method.

Assume that the federated system consists of K local clients. The local fault dataset of the k-th client is denoted as $D_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k}$, where C is the number of fault classes. The local diagnostic model is composed of a feature extractor $\phi_k(\cdot)$ and a classifier $g_k(\cdot)$, and the predicted probability is formulated as

$$z_{k,i} = g_k(\phi_k(x_{k,i}; \theta_k)), p_{k,i} = \text{softmax}(z_{k,i}) \quad (1)$$

where $z_{k,i}$ denotes the classification logits of the i-th sample, $p_{k,i}$ is the corresponding predicted probability, and θ_k represents the local model parameters.

Unlike methods that directly upload model parameters or raw data, each client only uploads diagnostic knowledge descriptors extracted from its local model. The federated center then uses these descriptors to update the global teacher model F_G and subsequently performs dynamic slimming through the gated student model F_S. The key distinction of the proposed framework is that relational knowledge is used throughout the whole process, including local knowledge description, server-side rare-fault-aware fusion, teacher model optimization, and gated student slimming. Therefore, rare-fault boundaries are preserved during both knowledge aggregation and lightweight compression.

Remark 1. *The proposed method keeps raw monitoring signals at local clients and only uploads compact knowledge descriptors, including softened outputs, class prototypes, relation matrices, and class-level statistics. This design reduces the exposure of raw vibration, acoustic, or*

current signals. Therefore, the proposed method should be understood as a raw-data-free and privacy-aware diagnostic framework rather than a method with a formal privacy guarantee. The upload cost of one client in each communication round is mainly determined by the descriptor size: $B_{rel} = b(N_a C + Cd + 2C^2 + 3C)$*, where b is the number of bytes of one floating-point value,* N_a *is the number of calibration samples, C is the number of classes, and d is the prototype dimension. The four terms correspond to output responses, class prototypes, two relation matrices, and class-level statistics, respectively. In comparison, parameter-based federated learning uploads model parameters with the cost* $B_{param} = b \mid \theta \mid$*. Since* N_a*, C, and d are usually much smaller than the number of local signal samples or model parameters, the proposed descriptor-based transmission is communication-efficient.*

3.2. Local Model Training and Multi-Granularity Diagnostic Knowledge Mining

In each federated iteration, the federated center first distributes the current global teacher model parameters to the local clients. The *k*-th client then updates its model using its local dataset. Considering that rare-fault classes may be insufficiently represented in local data, a class-balancing factor is introduced to regularize the local loss function:

$$\mathcal{L}_k^{\mathrm{ce}} = -\frac{1}{n_k}\sum_{i=1}^{n_k} \gamma_{k,y_{k,i}} \log p_{k,i}^{(y_{k,i})}, \quad \gamma_{k,c} = \frac{(n_{k,c}+\varepsilon)^{-\eta}}{\sum_{j=1}^{C}(n_{k,j}+\varepsilon)^{-\eta}}, \tag{2}$$

where $n_{k,c}$ denotes the number of samples belonging to the *c*-th fault class in the *k*-th client, ε is a small constant used to avoid division by zero, and η controls the compensation strength for rare classes. This design enables the local training stage to pay more attention to minority-fault classes, thereby preventing the subsequent federated fusion process from being dominated by majority-class samples. After model updating, the local clients do not upload raw monitoring signals. Instead, three types of diagnostic knowledge are extracted from the output layer, feature layer, and sensitivity-response layer, as shown in Figure 2.

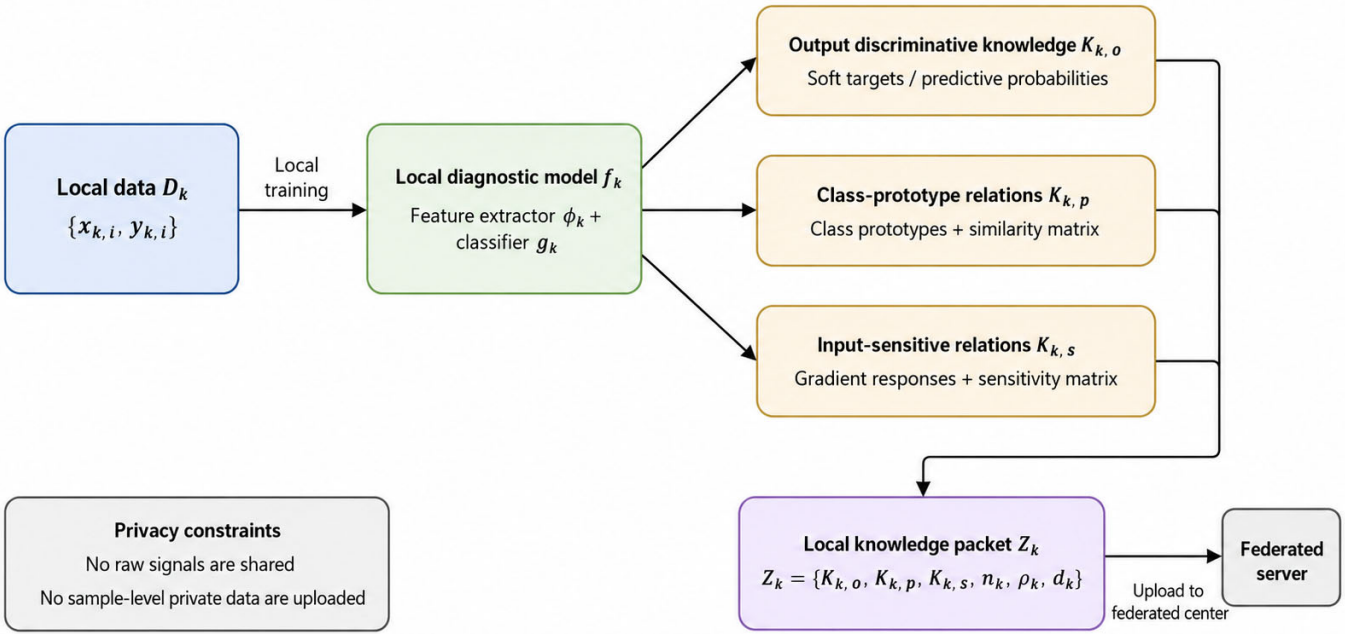

Figure 2. Local model training and multi-granularity diagnostic knowledge mining process.

First, output-discriminative knowledge is used to describe the soft decision boundaries of the model among different fault classes. The output distribution is softened using the temperature factor T, which is defined as

$$q_{k,i}^{T} = \mathrm{softmax}\left(\frac{z_{k,i}}{T}\right) \tag{3}$$

Compared with hard labels, $q_{k,i}^{T}$ preserves the similarity information between different fault classes and provides finer-grained discriminative constraints for the subsequent student model. Second, class-prototype relation knowledge is used to characterize the structural relationships among different fault classes in the feature space. Let $h_{k,i} = \phi_k(x_{k,i}; \theta_k)$ denote the deep feature extracted by the local model. The local class prototype of the c-th fault class is then defined as

$$\pi_{k,c} = \frac{1}{n_{k,c}} \sum_{i:y_{k,i}=c} \frac{h_{k,i}}{\|h_{k,i}\|_2} \tag{4}$$

The class-relation matrix is constructed based on the class prototypes as follows:

$$R_k^{p}(c, c') = \frac{\pi_{k,c}^{\top} \pi_{k,c'}}{\|\pi_{k,c}\|_2 \|\pi_{k,c'}\|_2} \tag{5}$$

This matrix reflects the relative distances and similarity relationships among different fault classes, preventing the student model from learning only sample-level outputs while ignoring the underlying fault-class structure during the slimming process. Finally, input-sensitive relation knowledge is introduced to describe the model response to weak fault features and boundary samples. For the c-th fault class, its sensitivity prototype is defined as

$$s_{k,c} = \frac{1}{n_{k,c}} \sum_{i:y_{k,i}=c} \frac{\nabla_{h_{k,i}} Z_{k,i}^{(c)}}{\|\nabla_{h_{k,i}} Z_{k,i}^{(c)}\|_2}, \tag{6}$$

The input-sensitive relation matrix is further constructed as

$$R_k^{S}(c, c') = \frac{s_{k,c}^{\top} s_{k,c'}}{\|s_{k,c}\|_2 \|s_{k,c'}\|_2} \tag{7}$$

Each component in the proposed pipeline is designed for a specific limitation in rare-fault-preserving federated slimming. Output knowledge preserves soft decision boundaries and global classification responses. Class-prototype relations preserve the structural distribution among fault classes, which is important for distinguishing rare faults from adjacent majority faults. Input-sensitive relations describe the response of the model to weak fault features and boundary samples, which helps retain rare-fault-sensitive channels. Rare-fault-aware weighting is used at the server to prevent reliable minority-fault knowledge from being suppressed by large-sample clients. Gated slimming further removes low-contribution channels while preserving relation-constrained diagnostic knowledge. Therefore, these components are complementary rather than simply stacked modules.

3.3. Rare-Fault-Aware Federated Relation Knowledge Fusion and Teacher Model Optimization

The knowledge uploaded by local clients may have different levels of reliability. Clients with large sample sizes may contribute substantial redundant knowledge and low-quality clients may produce low-confidence knowledge, whereas clients containing rare faults, despite having fewer samples, may carry critical class-boundary information. Therefore, the federated center constructs class-level knowledge weights according to sample scarcity, output confidence, and intra-class dispersion, as shown in Figure 3.

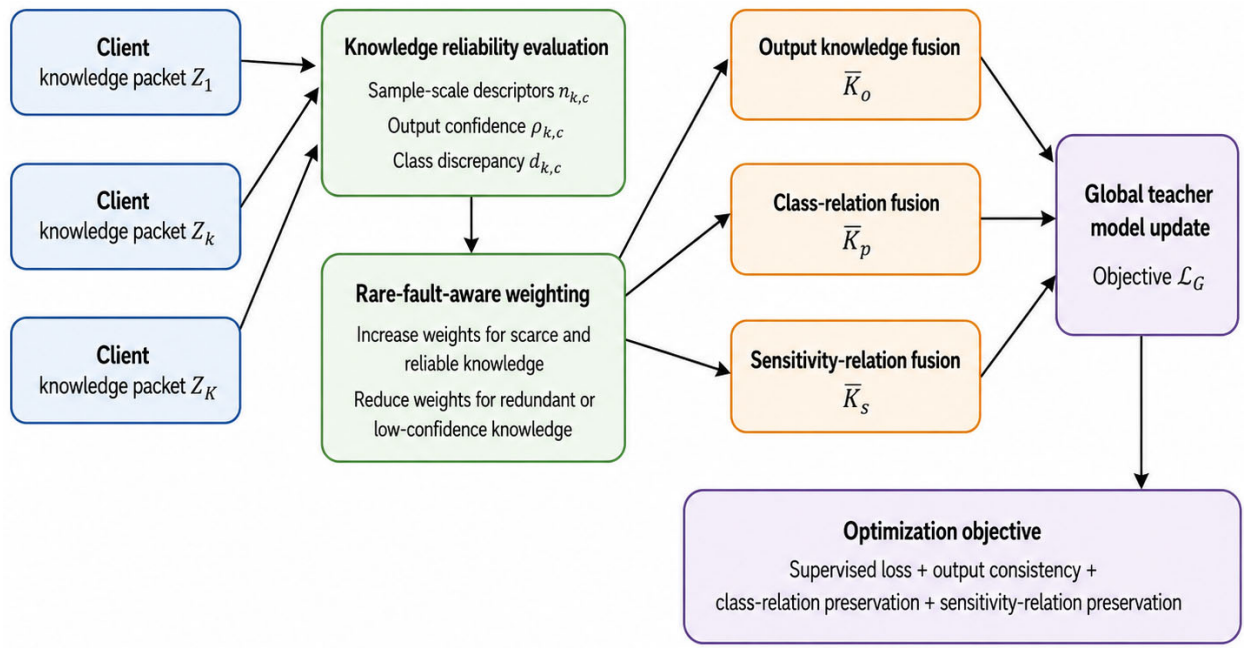

Figure 3. Process for federated relational knowledge fusion and teacher model optimization for rare-fault perception.

For the c-th fault class in the k-th client, its knowledge reliability is defined as

$$\rho_{k,c} = 1 - \frac{1}{n_{k,c} \log C} \sum_{i:y_{k,i}=c} H\left(q_{k,i}^{T}\right) \tag{8}$$

where $H(\cdot)$ denotes the information entropy. A larger $\rho_{k,c}$ indicates a more stable output distribution for the corresponding class. The intra-class dispersion is further defined as

$$d_{k,c} = \frac{1}{n_{k,c}} \sum_{i:y_{k,i}=c} \left\| \frac{h_{k,i}}{\|h_{k,i}\|_2} - \pi_{k,c} \right\|_2^2 \tag{9}$$

Based on this, the class-level federated knowledge weight is constructed as

$$\alpha_{k,c} = \frac{\exp(\beta_1 r_{k,c} + \beta_2 \rho_{k,c} - \beta_3 d_{k,c})}{\sum_{j \in \mathcal{K}_c} \exp(\beta_1 r_{j,c} + \beta_2 \rho_{j,c} - \beta_3 d_{j,c})}, \quad r_{k,c} = \frac{1}{\log(e + n_{k,c})} \tag{10}$$

where $\mathcal{K}_c$ denotes the set of clients containing knowledge of the c-th fault class; $r_{k,c}$ represents the class scarcity degree; and β_1, β_2, and β_3 control the effects of scarcity, reliability, and dispersion, respectively. The coefficients β_r, β_ρ, and β_d control the effects of class scarcity, output reliability, and intra-class dispersion, respectively. In the validation stage, β_r and β_ρ were selected from $\{0.5, 1.0, 1.5\}$ and β_d was selected from $\{0.25, 0.5, 1.0\}$. The final setting was $(\beta_r, \beta_\rho, \beta_d) = (1.0, 1.0, 0.5)$. This setting increases the contribution of reliable rare-fault knowledge while avoiding excessive amplification of unstable minority samples or excessive suppression of dispersed rare-fault features.

Based on $\alpha_{k,c}$, the federated center fuses the class prototypes and relation knowledge as follows:

$$\begin{gathered} \overline{\pi}_c = \sum_{k \in \mathcal{K}_c} \alpha_{k,c} \pi_{k,c} \\ \overline{R}^p(c,c') = \sum_{k \in \mathcal{K}_{c,c'}} \overline{\alpha}_{k,c,c'} R_k^p(c,c'), \overline{R}^s(c,c') = \sum_{k \in \mathcal{K}_{c,c'}} \overline{\alpha}_{k,c,c'} R_k^s(c,c') \end{gathered} \tag{11}$$

where $\mathcal{K}_{c,c'}$ denotes the set of clients that contain relation knowledge between class c and class c', and $\alpha_{k,c,c'}$ is obtained by normalizing $\alpha_{k,c}$ and $\alpha_{k,c'}$. The federated center then optimizes the global teacher model using the fused knowledge, with the objective function defined as

$$\begin{gathered}\mathcal{L}_G = \mathcal{L}_0^{ce} + \lambda_o T^2 \mathrm{KL}(\bar{q}^T \parallel q_G^T) + \lambda_p \|R_G^p - \overline{R}^p\|_F^2 + \lambda_s \|R_G^s - \overline{R}^s\|_F^2 \\ \theta_G^{t+1} = \theta_G^t - \eta_G \nabla_{\theta_G} \mathcal{L}_G\end{gathered} \tag{12}$$

where L_0^{ce} denotes the supervised loss computed on the public anchor set or a small calibration set at the federated center. If labeled anchor data are unavailable, this term can be removed, and the teacher model can be updated only through distillation and relation constraints. $\bar{q}^T$ represents the weighted fusion result of output-discriminative knowledge from multiple clients. R_G^p and R_G^s denote the class-prototype relations and input-sensitive relations obtained by the teacher model on the anchor samples, respectively. This process enables the global teacher model not only to inherit the classification capability of local clients, but also to preserve the fault-class structure and the sensitivity responses to weak fault features.

3.4. Redundancy-Gated Dynamic Slimming of the Student Model

The global teacher model integrates diagnostic knowledge from multiple clients, but it may also contain redundant knowledge. Redundant knowledge mainly refers to majority-class-biased responses, repetitive feature representations from similar clients, and low-contribution channels that have weak effects on rare-fault discrimination. If such information is directly transferred to the student model, it may increase model complexity and weaken rare-fault boundaries. Therefore, this study introduces learnable gates to suppress low-contribution channels, while relation constraints are used to retain channels associated with output discrimination, class-prototype structure, and input-sensitive rare-fault responses, as shown in Figure 4.

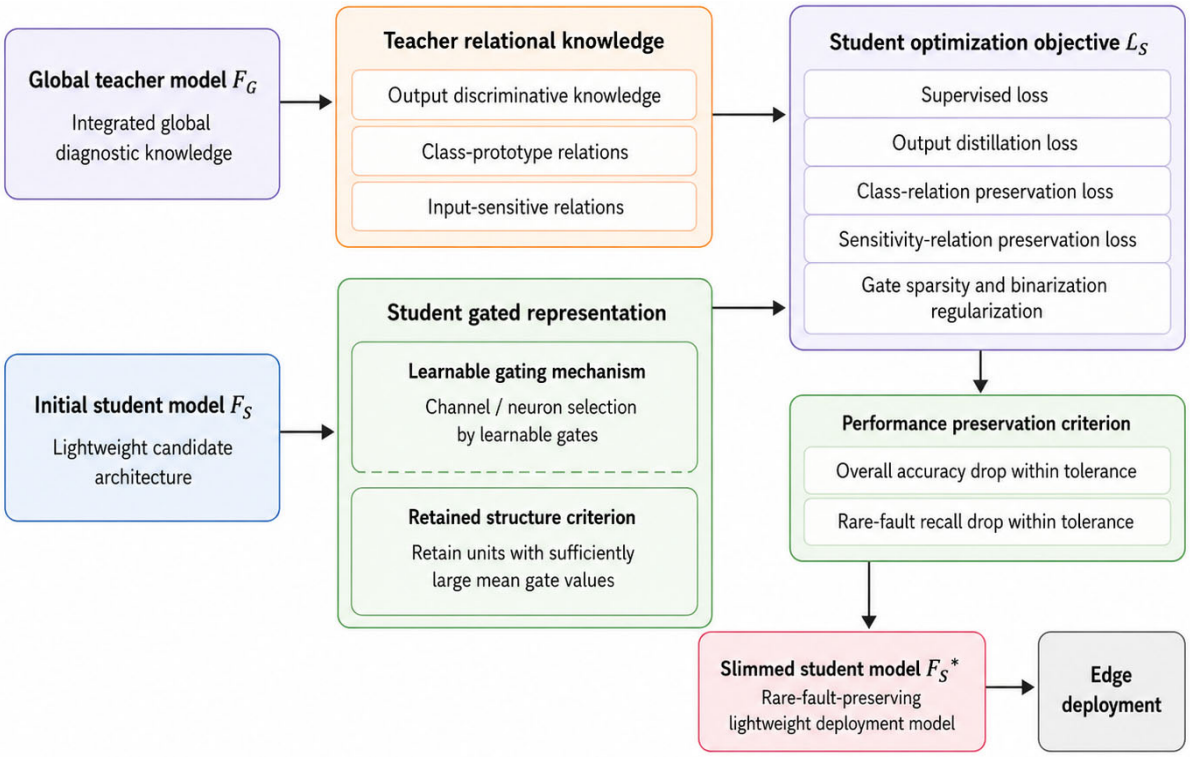

Figure 4. A dynamic slimming process for student models via redundant gating constraints.

Let $h_{S,l}$ denote the intermediate feature of the l-th layer in the student model. A learnable gating vector m_l is introduced to control the retention degree of channels or neurons in this layer:

$$m_l = \sigma(a_l), \ \widetilde{h}_{S,l} = m_l \odot h_{S,l} \tag{13}$$

where a_l denotes the gating parameter, $\sigma(\cdot)$ is the sigmoid function, and $\odot$ represents element-wise multiplication. The training objective of the student model consists of output distillation, class-relation preservation, sensitivity relation preservation, and gate sparsity regularization. A smaller gate value indicates that the corresponding channel contributes less to the relation-preserving objective and is more likely to carry redundant or weakly task-related information.

$$\begin{aligned} \mathcal{L}_S = \mathcal{L}_S^{ce} + \mu_o T^2 \mathrm{KL}(q_G^T \parallel q_S^T) + \mu_p \| R_G^p - R_S^p \|_F^2 + \mu_s \| R_G^s - R_S^s \|_F^2 + \\ \mu_g \sum_l \| m_l \|_1 + \mu_b \sum_l \| m_l (1 - m_l) \|_1 \end{aligned} \tag{14}$$

where the first four terms enable the student model to inherit the output-discriminative capability, class structure relations, and input-sensitive responses of the teacher model. The term $\sum_l \| m_l \|_1$ is used to compress redundant channels, while $\sum_l \| m_l(1 - m_l) \|_1$ encourages the gating values to converge toward either 0 or 1, thereby avoiding ambiguous structure selection during training. For the student objective in Equation (14), the output distillation weight is used as the reference term and is set to $\mu_o = 1.0$. The prototype relation and sensitivity relation weights are selected from $\{0.25, 0.5, 1.0\}$, and both are set to $\mu_p = \mu_s = 0.5$. This moderate setting preserves fault-class structures and weak-fault responses without over-constraining the student model. The sparsity and binary regularization weights are set to $\mu_g = 10^{-4}$ and $\mu_b = 10^{-3}$, respectively, to encourage compact and clear channel selection without causing excessive pruning or composite imbalance conditions.

After training, the retained structure is determined according to the mean gating value:

$$\mathcal{M}_S^* = \left\{ j \mid \vec{m}_{l,j} \geq \xi, \ j \in \mathcal{M}_S \right\} \tag{15}$$

where ξ is the structure-retention threshold, M_s^* denotes the retained structural set of the student model after gate-based dynamic slimming. To prevent excessive compression from degrading rare-fault recognition performance, a performance-preservation criterion is introduced:

$$\mathrm{Acc}(F_S^*) \geq \mathrm{Acc}(F_G) - \delta_a, \ \mathrm{Rec}_{rare}(F_S^*) \geq \mathrm{Rec}_{rare}(F_G) - \delta_r \tag{16}$$

where F_s^* represents the final slimmed student model constructed according to M_s^*, $\mathrm{Rec}_{rare}(\cdot)$ denotes the recall of rare-fault classes and δ_a and δ_r are the allowable degradation ranges of the overall accuracy and rare-fault recall, respectively. It should be noted that this criterion is used only during the offline slimming and model selection stage. In this stage, a small labeled calibration set, historical fault records, or laboratory-collected rare-fault samples can be used to prevent excessive compression from damaging rare-fault recognition. During online deployment, the slimmed model performs inference directly and does not require rare-fault labels.

3.5. Algorithm Implementation

The overall training and slimming procedure are summarized in Algorithm 1.

Algorithm 1. Relation knowledge-guided federated dynamic slimming

Input: Local datasets D_k, calibration set D_a, global teacher model F_G, student model F_S, maximum communication rounds R.
Output: Slimmed student model F_S^*.
Initialize the global teacher model F_G, local client models F_k, student model F_S, and gate parameters.
For each communication round $r = 1, 2, \ldots, R$:
The server broadcasts the current global teacher model F_G to all clients.
For each client k:
Update the local model F_k on D_k using the class-balanced loss.
Extract output knowledge Q_k, class-prototype relation R_k^p, and input-sensitive relation R_k^s.
Calculate the class sample number $n_{k,c}$, output reliability $\rho_{k,c}$, and intra-class dispersion $d_{k,c}$.
Upload the knowledge descriptor

$$Z_k = \left\{ Q_k, R_k^p, R_k^s, n_{k,c}, \rho_{k,c}, d_{k,c} \right\}$$

to the server.
End for.
The server calculates the rare-fault-aware fusion weight $\alpha_{k,c}$.
Fuse the multi-client output knowledge, class-prototype relations, and input-sensitive relations.
Update the global teacher model F_G on the calibration set D_a.
Stop teacher training if the stopping criterion is satisfied.
End for.
Train the gated student model F_S under the guidance of F_G.
Update the student model parameters and gate parameters by back-propagation.
Calculate the mean gate value of each channel.
Retain the channels whose mean gate values are not smaller than the retention threshold.
Construct the slimmed student model F_S^*.
If F_S^* satisfies the accuracy and rare-fault recall preservation criterion, output F_S^*.
Otherwise, relax the retention threshold and fine-tune the student model.

In Algorithm 1, the calibration set D_a is only used for server-side teacher optimization and is not included in local client training or final testing.

4. Experimental Results and Analysis

The effectiveness of the proposed method is evaluated on two public datasets. The CWRU bearing fault dataset is used to assess class absence, low-quality small-sample conditions, quality–scale imbalance, and sample–scale imbalance under single-modal vibration signals. The HUST motor multi-modal motor fault dataset contains both vibration and acoustic signals and is used to evaluate multi-modal fragmentation, cross-condition generalization, and modality missing at the deployment stage. These two datasets are complementary in terms of signal modality, fault pattern, and experimental setting.

4.1. Experimental Setup and Comparison Methods

The CWRU dataset is provided by the Case Western Reserve University Bearing Data Center [33], and its experimental platform is shown in Figure 5. In this study, the data collected under a 1 HP load, a rotational speed of 1797 r/min, and a sampling frequency of 48 kHz are selected to construct the diagnosis task. Signals of each class are segmented into samples using a fixed-length window and then normalized by zero-mean standardization.

To simulate multi-client federated diagnosis, the dataset is divided among three clients, and three scenarios are designed, namely, composite imbalance, quality–scale imbalance, and sample–scale imbalance, and the experimental setup is shown in Table 1.

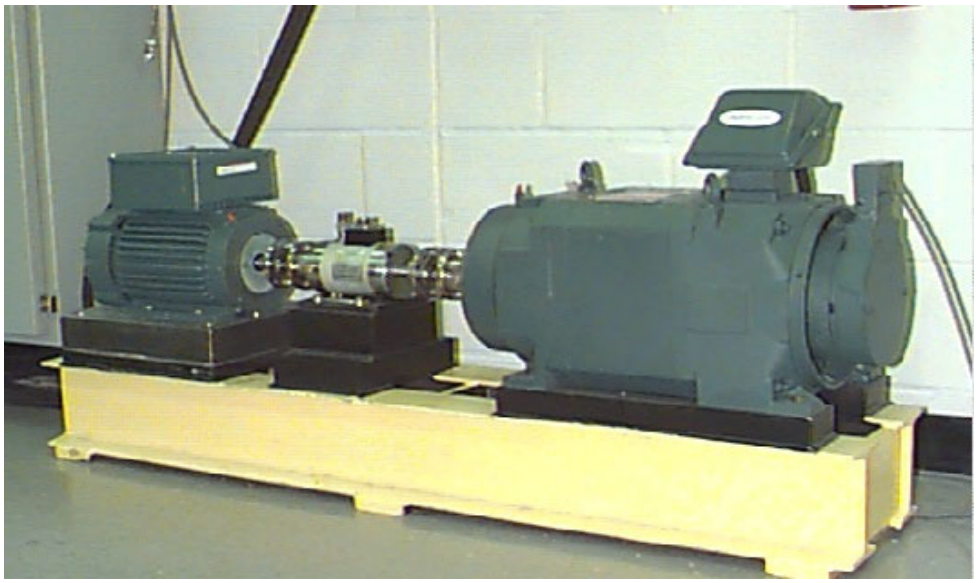

Figure 5. Experimental platform of the CWRU motor bearing fault dataset [33].

Table 1. Federated imbalance scenario design for the CWRU dataset.

Scenario	Client 1	Client 2	Client 3	Validation Objective
Composite imbalance	20 × 10, 2 dB noise, with OR14	200 × 9, without OR14	200 × 9, without OR14	Verify rare-fault preservation under class absence, low-quality small samples, and majority-class dominance
Quality–scale imbalance	20 × 10, 2 dB noise	200 × 10, no additional noise	200 × 10, no additional noise	Verify reliable utilization of noisy small-sample knowledge
Sample–scale imbalance	20 × 10, no additional noise	40 × 10, no additional noise	200 × 10, no additional noise	Verify suppression of redundant knowledge from large-sample clients

The HUST motor dataset is a public multi-modal motor fault dataset, and its experimental platform is shown in Figure 6 [34]. The platform was built based on the Spectra Quest Mechanical Fault Simulator, and the monitored object is a motor system under different health conditions. The dataset contains six operating states, namely, healthy condition, bearing fault, rotor bow, broken rotor bar, rotor misalignment, and voltage unbalance. The operating frequencies are 5, 10, 20, and 30 Hz, and the sampling frequency is 25.6 kHz. Each file contains 163,840 data points. In this study, the dataset is used to construct three experimental scenarios: modality-fragmented rare-fault diagnosis, cross-condition generalization, and modality missing at the deployment stage, and the experimental setup is shown in Table 2.

The details of the network structure and parameter settings under different datasets during the experiment are shown in Table 3.

The CWRU experiments are designed to highlight single-modal class imbalance and the influence of model compression. Therefore, in addition to general federated baselines, quantization, pruning, and knowledge distillation methods related to model compression and communication efficiency are introduced. The HUST motor experiments focus on

multi-modal and cross-condition heterogeneous scenarios. Accordingly, FedBN, FedDF, and FedSlim are further included on the basis of general federated baselines, since these methods are able to address distribution shift, server-side ensemble distillation, and model-structure differences, as shown in Table 4.

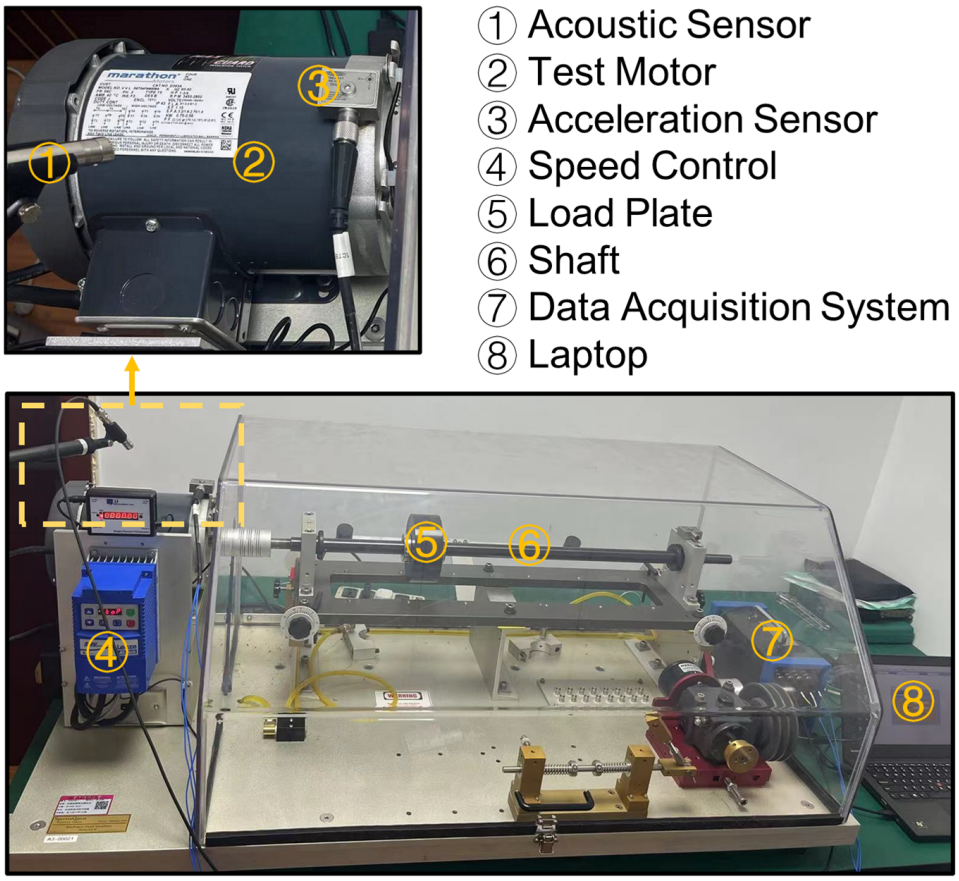

Figure 6. Experimental platform of the HUST motor multi-modal motor fault dataset [34].

Table 2. Multi-modal federated experimental scenario design for the HUST motor dataset.

Scenario	Training Setting	Testing Setting	Validation Objective
Modality-fragmented rare fault	Client 1: 5/10 Hz vibration modality, all classes included but with few BRB samples; Client 2: 20 Hz acoustic modality, without BRB; Client 3: 30 Hz dual modalities, with few BRB samples	Mixed-condition test set	Verify rare-fault knowledge preservation under modality fragmentation
Cross-condition generalization	Clients 1, 2, and 3 are trained using 5, 10, and 20 Hz data, respectively, with both modalities included	30 Hz unseen-condition test set	Verify the generalization ability of relation knowledge under operating-frequency variation
Modality missing at deployment	Multi-clientel dual-modal data are used during training	Vibration-only, acoustic-only, and random modality-missing tests	Verify the edge-deployment robustness of the gated student model

Table 3. Main implementation settings of the proposed method.

Item	Setting
Calibration set	CWRU uses 5 labeled samples per class, and HUST motor uses 10 labeled samples per class. These samples are excluded from local training and final testing.
Client update	Three clients participate in each round. Each client performs 5 local epochs before uploading knowledge descriptors.
Model architecture	The teacher uses a 1D-CNN with three Conv-BN-ReLU-Pool blocks. The student uses a narrower 1D-CNN with learnable channel gates after convolution blocks. For the HUST motor, vibration and acoustic signals are processed by two parallel branches and then fused.
Optimization	The Adam optimizer is used. The learning rate is 1×10^{-3} for local and teacher models, 5×10^{-4} for the student model, and 1×10^{-3} for gate parameters. The batch size is 64.
Hyperparameters	$T = 3$, $\eta = 0.5$, $\epsilon = 10^{-6}$, $\beta_r = 1.0$, $\beta_\rho = 1.0$, $\beta_d = 0.5$. The weights of output, prototype, and sensitivity losses are set to 1.0, 0.5, and 0.5, respectively.
Slimming setting	Gate parameters are updated by back-propagation. The retention threshold is $\xi = 0.5$. If the performance criterion is not satisfied, ξ is reduced by 0.05 and the student model is fine-tuned.
Model training	The maximum number of communication rounds is 100. Teacher training stops if calibration Macro-F1 does not improve for 10 rounds. Student training lasts 50 epochs at most and stops early if validation loss does not improve for 10 epochs.

Table 4. Detailed description of comparison methods.

Method	Method Category	Brief Description	Dataset Use
DNN	Single-client baseline	Local deep diagnostic model without federated collaboration.	CWRU; HUST motor
Pruning	Model compression	Removes low-contribution weights, channels, or neurons.	CWRU
FedAvg	Classical FL	Aggregates local model parameters at the server.	CWRU; HUST motor
FedProx [35]	Federated optimization	Adds a proximal term to mitigate client drift.	CWRU; HUST motor
FedProto [36]	Federated representation	Aggregates class prototypes for cross-client structure alignment.	CWRU; HUST motor
Fed-KD [37]	Federated distillation	Uses knowledge distillation for communication-efficient FL.	CWRU; HUST motor
Fed-Pruning [38]	Federated compression	Introduces pruning into federated training.	CWRU
FedBN [39]	Cross-domain FL	Keeps local BN statistics to alleviate feature shift.	HUST motor
FedDF [40]	Federated distillation	Fuses heterogeneous client models through server-side ensemble distillation.	HUST motor
FedSlim [41]	Federated slimming	Applies channel-level slimming or structural compression in FL.	HUST motor
Proposed	Proposed method	Rare-fault-aware weighting, relation knowledge fusion, and gated dynamic slimming.	CWRU; HUST motor

To ensure fairness, all compared methods used the same preprocessing procedure, input length, training/testing split, and 1D-CNN backbone whenever applicable.

Compression-based methods were compared under comparable parameter budgets. Hyperparameters were selected only on the validation set, with the same search budget used for all methods. Each experiment was repeated five times with different random seeds and regenerated client partitions. Statistical significance between the proposed method and the strongest baseline was evaluated using a paired t-test, with $p < 0.05$ considered significant.

The evaluation metrics include average accuracy, Macro-F1, rare-class recall, the number of parameters, FLOPs, inference time, and GPU memory usage. In the CWRU dataset, OR14 is defined as the primary rare-fault class, while in the HUST motor dataset, the broken rotor bar fault, BRB, is defined as the primary rare-fault class. Macro-F1 is used to prevent the performance of minority classes from being masked by majority classes, and rare-class recall is used to directly evaluate the core objective of the proposed method.

4.2. Results Under the CWRU Composite Imbalance Scenario and Rare-Fault Recognition Analysis

The composite imbalance scenario is the most challenging setting in the CWRU experiments. Since OR14 is provided only by Client 1, which contains a limited number of samples and is contaminated by noise, conventional federated aggregation tends to regard its information as weak local knowledge or low-confidence knowledge, causing it to be overwhelmed by the information from the majority of clients. Table 5 and Figure 7 present the quantitative results under this scenario.

Table 5. Diagnostic results for the CWRU compound imbalance scenario.

Methods	Client 1 Acc/%	Client 2 Acc/%	Client 3 Acc/%	Average Acc/%	Macro-F1/%	OR14 Recall/%
DNN	49.8	53.4	55.8	53.0	50.6	18.0
Pruning	49.0	53.6	55.0	52.5	49.7	20.0
FedAvg	62.4	64.6	65.0	64.0	60.8	34.0
Fed-Prox	66.8	69.4	70.2	68.8	65.2	46.0
Fed-Pruning	68.4	70.0	70.6	69.7	66.3	48.0
Fed-Proto	71.6	73.2	74.0	72.9	70.5	56.0
Fed-KD	73.2	75.0	75.8	74.7	72.0	58.0
Proposed	83.6	85.2	86.0	84.9	84.0	80.0

As shown in Table 5, conventional single-client models and standard compression methods exhibit poor performance on Client 1, indicating that a small sample size and low signal-to-noise ratio weaken the feature learning capability of local models. FedAvg improves the average accuracy to 64.0%, but its OR14 recall remains only 34.0%, suggesting that parameter averaging is easily dominated by clients with larger sample sizes and more complete class distributions. Fed-Prox alleviates client drift through a proximal constraint and increases the average accuracy to 68.8%; however, its optimization objective does not explicitly emphasize rare-class preservation. Fed-Proto mitigates partial inter-class structural shift by using class prototypes, increasing the OR14 recall to 56.0%, but it lacks input-sensitive relations and gated redundancy suppression. Fed-KD further enhances global consistency by exploiting soft output knowledge, yet its OR14 recall remains 58.0%. In contrast, the proposed method achieves an average accuracy, Macro-F1, and OR14 recall of 84.9%, 84.0%, and 80.0%, respectively, demonstrating that multi-granularity relation knowledge and rare-fault-aware weighting can significantly improve minority-class boundary preservation.

Combining Table 5 and the confusion matrices, the differences between the methods are reflected not only in the average accuracy results but more importantly in the misclassification of the rare OR14 class. DNN, quantization, and pruning can recognize several majority fault classes, but they show severe misclassification for OR14, which is mainly confused with the adjacent outer-race faults OR07 and OR21. This indicates that single-client

training and conventional compression methods struggle to preserve rare-fault boundaries under small-sample and low-SNR conditions. FedAvg strengthens the overall diagonal pattern and improves the average accuracy to 64.0%, but its OR14 recall remains only 34.0%, suggesting that conventional parameter aggregation is easily dominated by majority clients and majority classes. Fed-Prox and Fed-Proto improve the results through proximal regularization and class-prototype alignment, respectively, but evident confusion still exists between OR07, OR14, and OR21. Fed-KD further improves global consistency using soft output knowledge, yet its OR14 recall is only 58.0%, showing that output distillation alone is insufficient to preserve fine-grained boundaries between similar faults.

In contrast, the proposed method achieves an average accuracy of 84.9%, a Macro-F1 of 84.0%, and an OR14 recall of 80.0%, while substantially reducing the misclassification between OR14 and its adjacent classes OR07 and OR21. This demonstrates that class-prototype relations preserve the structural distribution of similar faults, input-sensitive relations enhance weak-response boundaries, rare-fault-aware weighting increases the contribution of reliable OR14 knowledge from the small-sample client, and the gated student model retains key diagnostic channels after compression. These results confirm that the proposed method can effectively preserve rare-fault recognition capability under composite imbalance conditions.

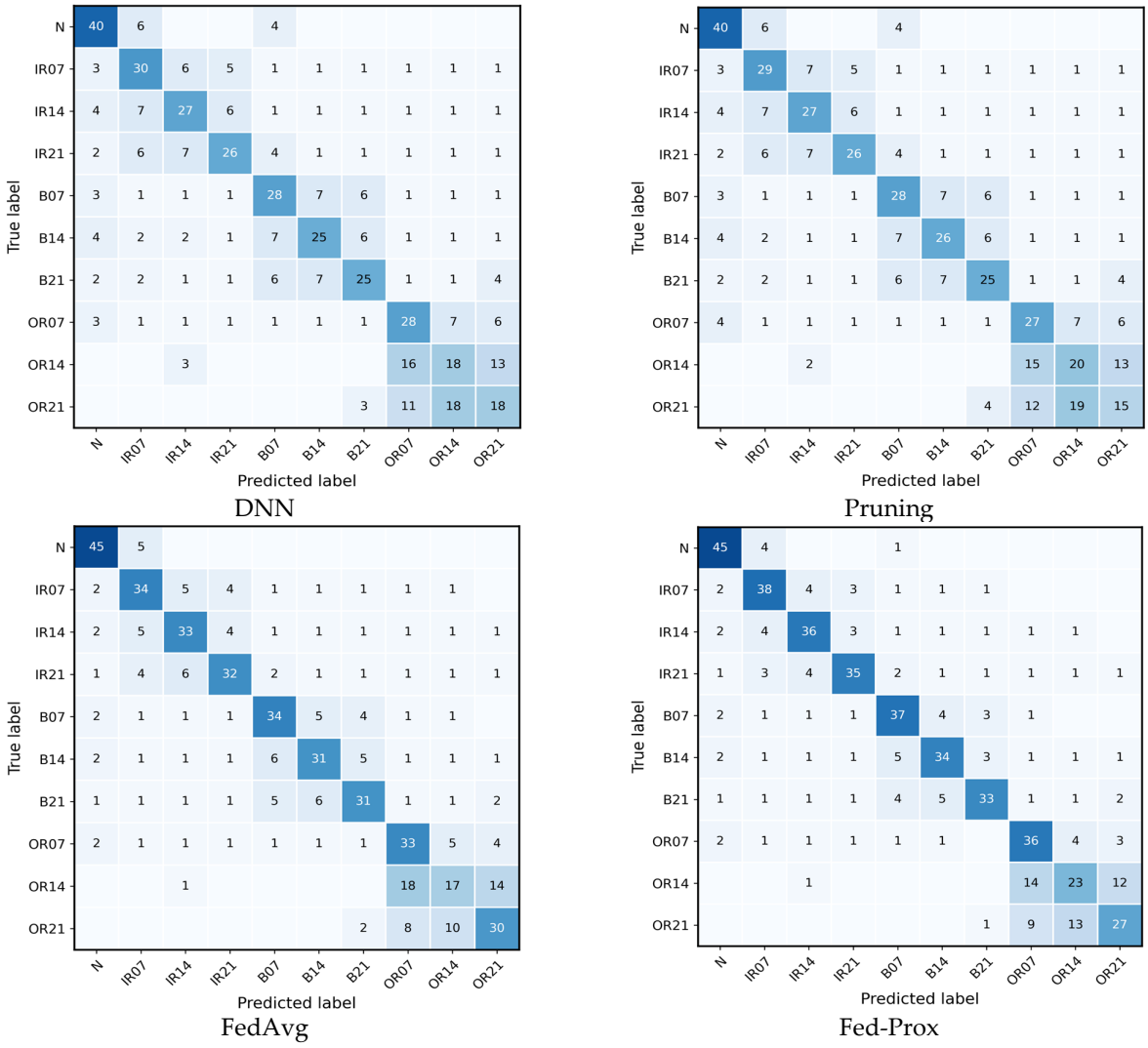

Figure 7. *Cont.*

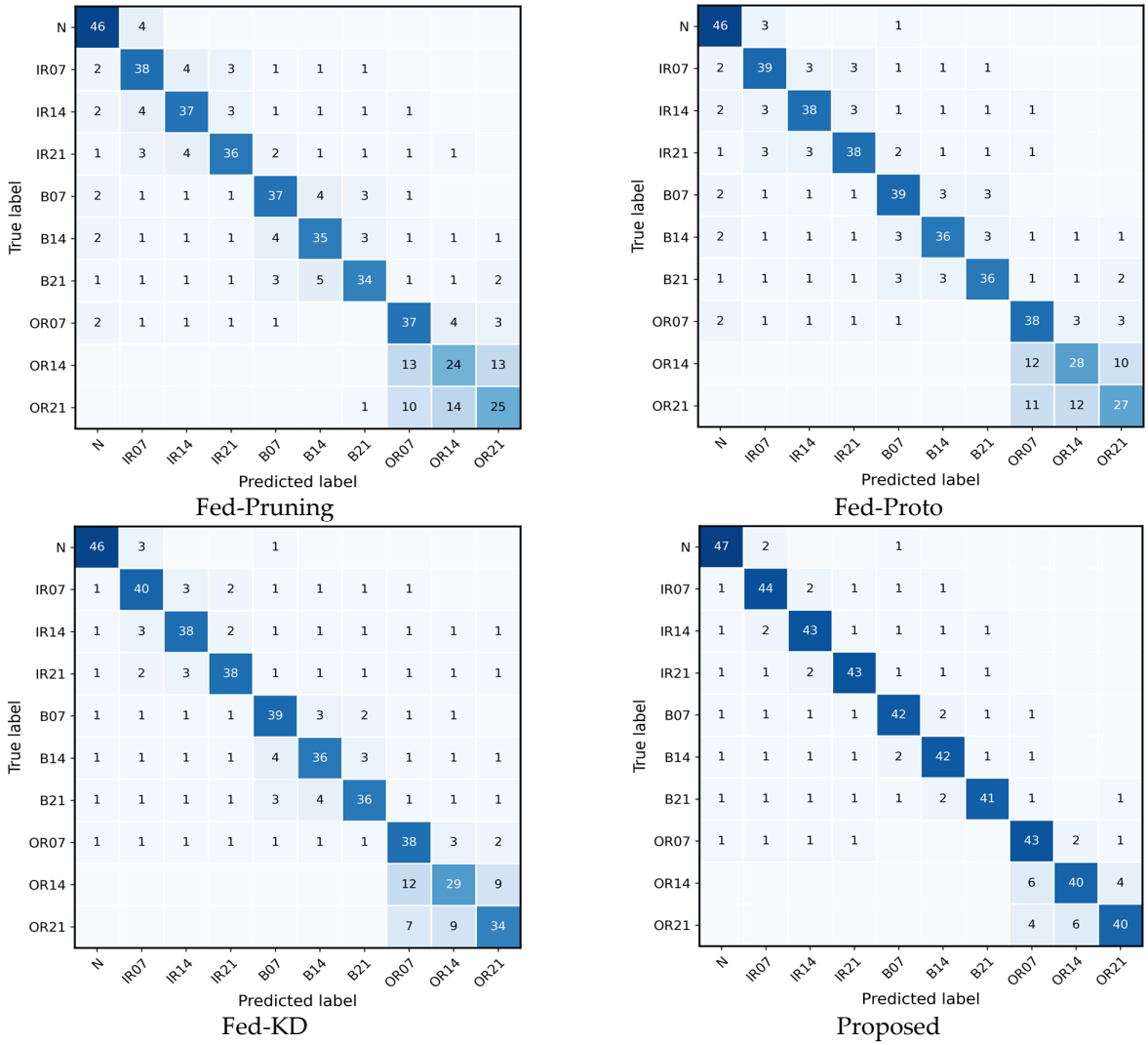

Figure 7. Confusion matrices of different methods under the CWRU composite imbalance scenario.

To further visualize the effect of input-sensitive relation knowledge, feature-response heatmaps were generated for rare OR14 test samples. The response intensity was calculated from the channel-wise feature activation weighted by the sensitivity of the rare-fault output to the corresponding feature channel and then normalized to $[0, 1]$. Fed-KD, the proposed method without input-sensitive relation, and the complete proposed method were compared, as shown in Figure 8.

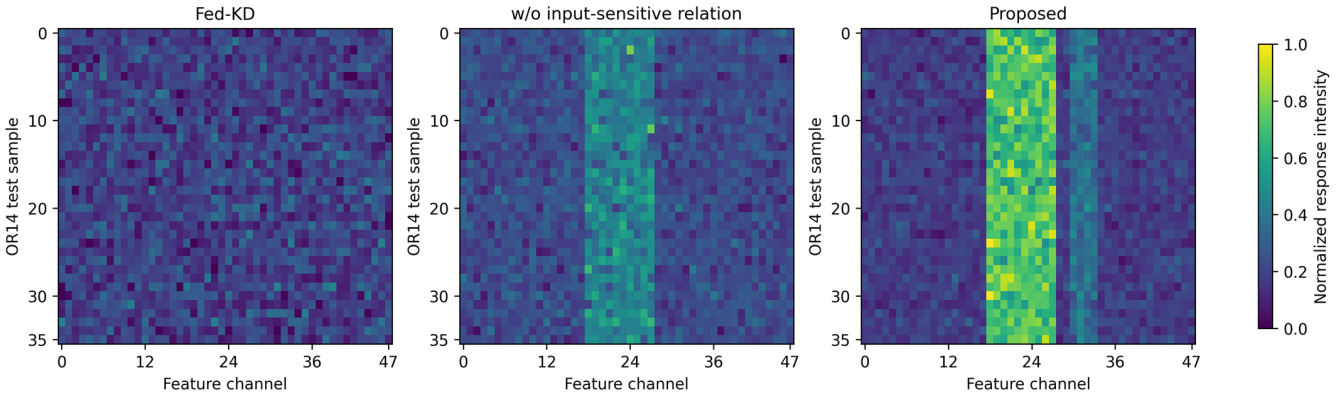

Figure 8. Feature-response heatmaps of different methods for rare OR14 samples.

In Figure 8, Fed-KD produces scattered feature responses, indicating that output-level distillation alone cannot clearly focus on weak rare-fault-related channels. After removing input-sensitive relation knowledge, the response region becomes clearer than Fed-KD, but the activation is still not sufficiently concentrated. In contrast, the complete proposed method produces stronger and more stable responses on a group of rare-fault-sensitive channels. This indicates that input-sensitive relation knowledge helps the student model retain weak-response patterns associated with OR14 and reduces the confusion between OR14 and adjacent outer-race faults.

4.3. Robustness and Deployment Efficiency Analysis Under CWRU Scenarios

To further verify the stability of the proposed method under different imbalance intensities, Figure 9 presents radar chart comparisons under three scenarios: composite imbalance, quality–scale imbalance, and sample–scale imbalance. Each subplot includes four metrics: Client 1 accuracy, average accuracy, Macro-F1, and OR14 recall. To ensure figure readability, only six representative methods are shown, namely, DNN, FedAvg, Fed-Prox, Fed-Proto, Fed-KD, and the proposed method.

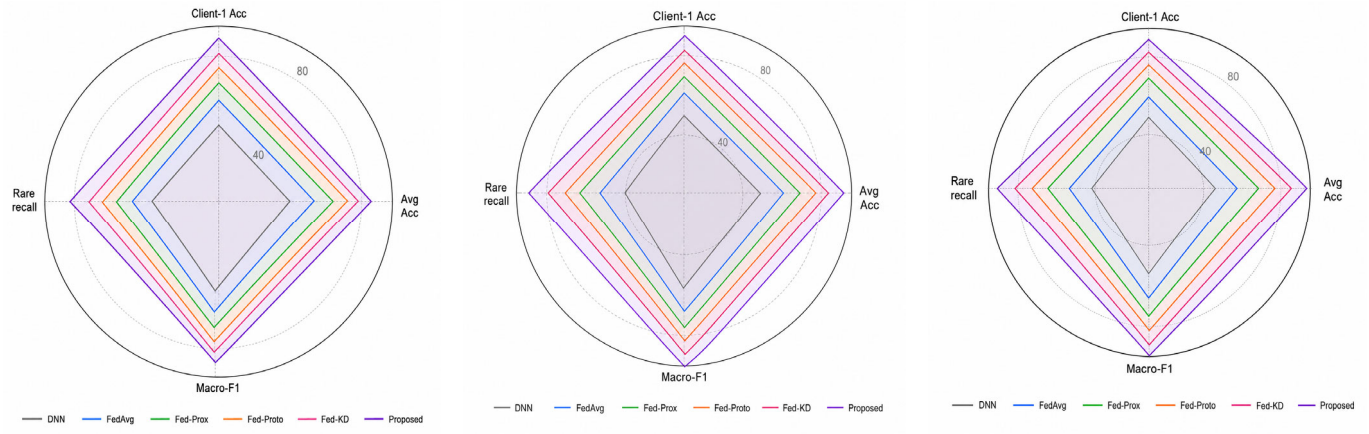

Figure 9. Radar chart comparison of representative methods under three imbalanced scenarios on the CWRU dataset.

As shown in Figure 8, the overall performance of all methods improves as the degree of imbalance decreases, while the proposed method maintains the largest radar coverage area across all three scenarios. In the compound imbalance scenario, the advantage is most evident in terms of OR14 recall, indicating that the proposed method can effectively address the rare-fault suppression problem caused by the coexistence of missing fault categories and low-quality few-shot samples. In the quality–scale imbalance scenario, the proposed method still achieves superior performance in Client 1 accuracy and Macro-F1, demonstrating that the knowledge reliability evaluation can select effective knowledge from noisy few-shot clients. In the scale imbalance scenario, the proposed method continues to outperform Fed-Proto and Fed-KD, suggesting that the gated student model is not only used for parameter compression but also capable of suppressing the propagation of repetitive knowledge from large-sample clients.

To make the deployment efficiency comparison reproducible, all models used the same input length, preprocessing procedure, and batch size. The inference latency was measured with a batch size of 1 to simulate online diagnosis. Each model was first executed for 100 runs and then evaluated over 1000 repeated inference runs. The reported latency is

the average single-sample inference time. FLOPs and parameter numbers were calculated using the same profiling tool; the specific settings are shown in Table 6.

Table 6. Method hardware deployment setup.

Item	Setting
Hardware	NVIDIA RTX 3060Ti GPU and Intel i7 CPU
Software	Python 3.9, PyTorch 1.13, CUDA 11.7
Batch size	1
Pre-training	100
Repeated runs	1000
Time delay	Average single-sample inference time
Memory metric	Peak allocated inference memory

Under the above deployment setup, Figure 10 shows that FedAvg and Fed-Prox require a higher computational cost and GPU memory because they retain the complete global model structure.

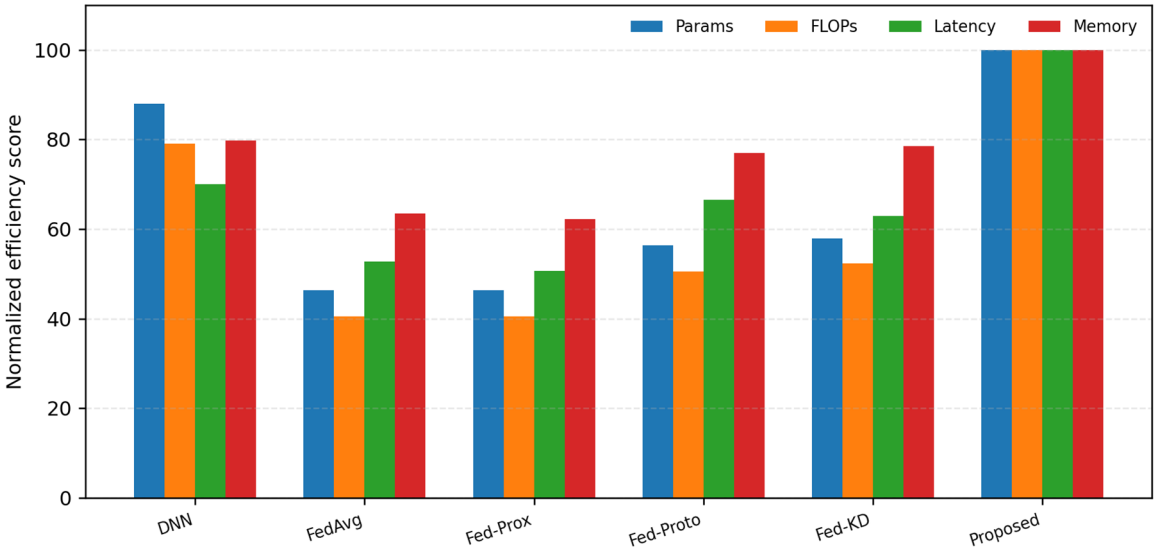

Figure 10. Deployment efficiency comparison of representative methods on the CWRU dataset.

Fed-Proto and Fed-KD outperform FedAvg in knowledge transfer, but their compression capability remains limited. In contrast, the proposed method selectively preserves key channels and parameters through the gated student network, thereby reducing redundant computation while retaining relation knowledge. Consequently, it achieves the highest overall deployment efficiency. The CWRU results demonstrate that the proposed method can simultaneously improve rare-fault recall and deployment-oriented efficiency under the tested hardware and measurement protocol.

4.4. Multi-Modal Federated Diagnosis Results on HUST Motor

To verify the applicability of the proposed method under multi-modal and cross-condition settings, further experiments are conducted on the HUST motor dataset. The main challenge of this dataset is not simple sample–scale imbalance but modality fragmentation, operating-condition variation, and modality missing at the deployment stage. Therefore, FedBN and FedDF are additionally included as baseline methods in this section, and BRB is defined as the primary rare-fault class. Table 7 presents the quantitative results under the modality-fragmented rare-fault scenario.

Table 7. Diagnostic results under the modality-fragmented rare-fault scenario on the HUST motor dataset.

Methods	Client 1 Acc/%	Client 2 Acc/%	Client 3 Acc/%	Average Acc/%	Macro-F1/%	BRB Recall/%
DNN	61.2	63.4	63.8	62.8	60.5	35.0
FedAvg	69.4	70.8	71.2	70.5	68.1	46.0
FedProx	72.6	73.9	74.3	73.6	71.8	50.0
FedBN	74.2	75.3	75.8	75.1	73.6	52.0
FedProto	77.1	78.6	79.2	78.3	76.9	61.0
FedDF	78.5	79.8	80.4	79.6	78.2	63.0
Fed-KD	80.0	81.5	82.1	81.2	80.0	66.0
FedSlim	79.3	80.9	81.2	80.5	78.7	64.0
Proposed	87.2	88.8	89.2	88.4	87.8	84.0

In Table 7, Local-DNN achieves an average accuracy of only 62.8% and a BRB recall of 35.0%, indicating that a single-client model struggles to learn multi-modal information and rare-fault features simultaneously. FedAvg improves the overall performance, but its BRB recall remains limited, suggesting that simple parameter aggregation is easily dominated by majority modalities and majority classes. FedBN alleviates cross-condition and cross-modal feature shifts by retaining local normalization statistics, while FedProto improves structural consistency through class prototypes; both methods outperform FedAvg. FedDF and Fed-KD further enhance global consistency using distillation knowledge, but they mainly rely on output responses or ensemble outputs, making it difficult to fully characterize inter-modal class structures and input-sensitive responses. The proposed method achieves 88.4% average accuracy, 87.8% Macro-F1, and 84.0% BRB recall, outperforming Fed-KD by 7.2, 7.8, and 18.0 percentage points, respectively. This indicates that the proposed method can preserve rare-fault knowledge under modality-fragmented and client-heterogeneous conditions.

Figure 11 shows that Local-DNN and FedAvg can identify majority classes such as H and BF to some extent, but they suffer from severe misclassification for the BRB class, which is mainly confused with BF and UNBAL. This phenomenon is consistent with the mechanism of broken rotor bar faults: a broken rotor bar induces electromagnetic torque fluctuations, whose responses are similar to those caused by voltage unbalance, and it may also trigger mechanical vibration variations. FedBN and FedProto reduce part of the cross-modal confusion, but evident overlap between BRB and UNBAL remains. Fed-KD improves the overall diagonal structure, yet the rare-class boundary is still not sufficiently clear. In contrast, the proposed method presents the clearest diagonal structure, with a substantially increased number of correctly identified BRB samples and significantly reduced misclassification among adjacent classes. This indicates that relation knowledge fusion improves boundary separability among similar fault categories.

4.5. Cross-Condition Generalization Results on the HUST Motor Dataset

The 5, 10, and 20 Hz operating conditions are used as local training conditions, while the 30 Hz condition is treated as an unseen testing condition. This setting is designed to examine whether federated knowledge fusion can learn fault-discriminative relations that remain stable across different operating frequencies. Table 8 presents the cross-condition generalization results.

Cross-condition testing significantly degrades the performance of all methods, indicating that operating-frequency variation leads to feature distribution shifts. FedBN outperforms FedProx in this scenario, suggesting that retaining local normalization statistics helps alleviate feature shifts across different operating conditions. FedDF and Fed-KD achieve average accuracies of 75.6% and 76.8%, respectively, showing that distillation

mechanisms can improve global consistency under unseen conditions. The proposed method achieves 84.3% average accuracy, 83.5% Macro-F1, and 78.0% BRB recall, while Fed-KD achieves 76.8% average accuracy, 75.1% Macro-F1, and 61.0% BRB recall. These results demonstrate that the proposed method remains effective under cross-condition and cross-device-like federated heterogeneity.

Figure 12 further illustrates the feature distributions on the unseen-condition test set. FedAvg shows evident overlap among different fault classes, especially with blurred boundaries among BF, BRB, and UNBAL. FedProto improves the structure of class centers, but the intra-class dispersion remains relatively large. Fed-KD enhances class compactness through output knowledge distillation, yet overlap between BRB and UNBAL still exists. In contrast, the proposed method forms more compact intra-class distributions and clearer inter-class separations, indicating that multi-granularity relation knowledge improves the structural consistency of the feature space and enables the model to maintain strong fault discriminability under unseen operating frequencies.

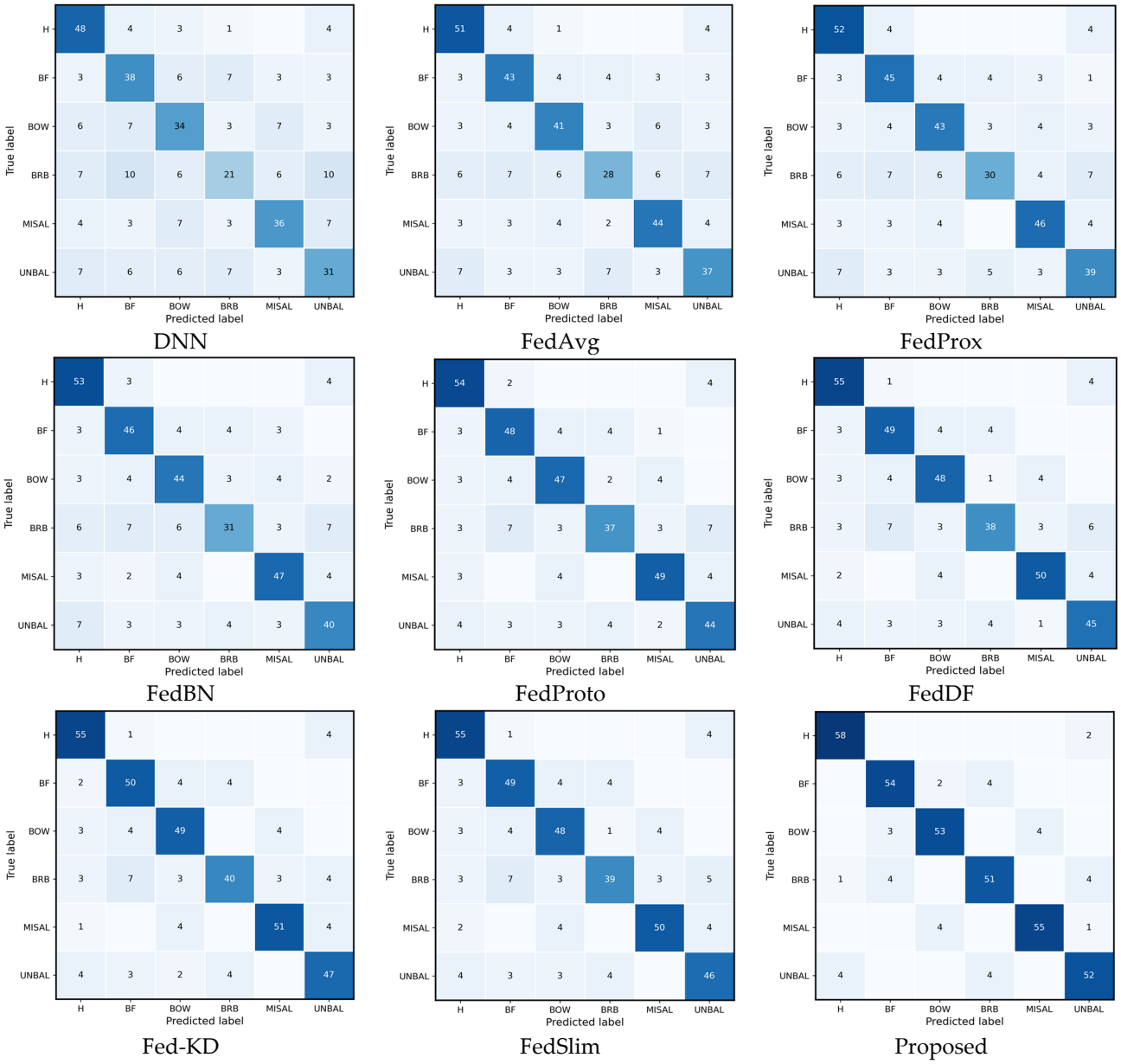

Figure 11. Confusion matrices of all methods for multi-modal federated diagnosis on the HUST motor dataset.

Table 8. Diagnostic results under the cross-condition generalization scenario on the HUST motor dataset.

Methods	Avg Acc/%	Macro-F1/%	BRB Recall/%
Local-DNN	58.9	55.8	30.0
FedAvg	65.7	63.2	40.0
FedProx	68.2	66.0	43.0
FedBN	70.4	68.5	46.0
FedProto	72.8	71.0	55.0
FedDF	75.6	73.9	58.0
Fed-KD	76.8	75.1	61.0
FedSlim	75.0	72.8	59.0
Proposed	84.3	83.5	78.0

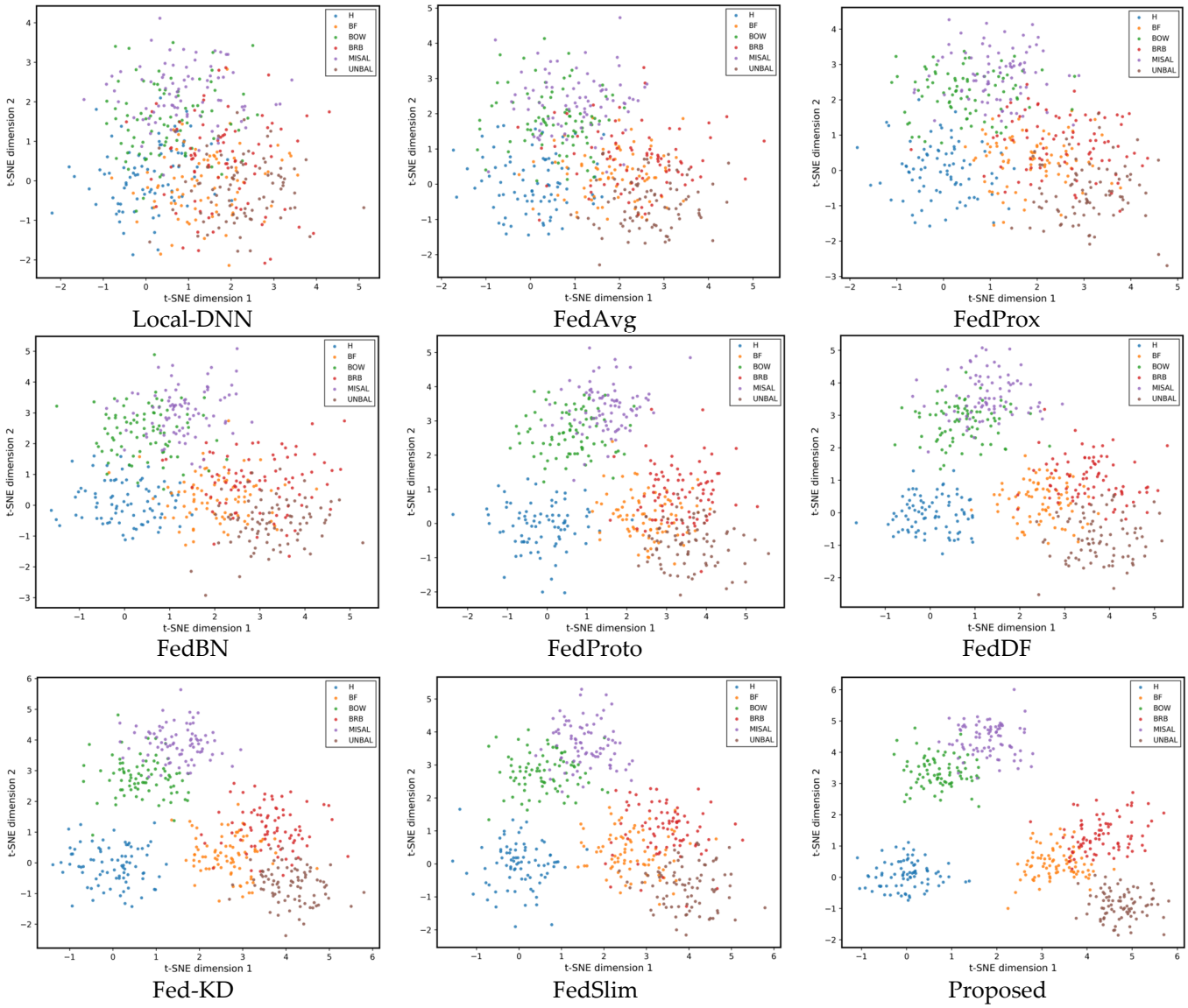

Figure 12. t-SNE visualization of features extracted by different methods under the cross-condition generalization scenario on the HUST motor dataset.

In practical deployment, the edge side may only have access to a single-modal signal due to sensor failure, communication constraints, or installation limitations. Therefore, the diagnostic capability of the model is further evaluated under vibration-only, acoustic-only,

and random modality-missing conditions. This scenario is designed to verify whether the gated dynamic slimming mechanism of the student model can preserve effective diagnostic knowledge during lightweight deployment, as shown in Table 9.

Table 9. Diagnostic results under the modality-missing deployment scenario on the HUST motor dataset.

Methods	Single Vibration Acc/%	Acoustic Acc/%	Random Modal Dropout Acc/%	BRB Recall/%
Local-DNN	65.1	61.3	58.8	31.0
FedAvg	72.3	68.6	69.4	43.0
FedProx	74.5	70.2	71.6	47.0
FedBN	76.2	72.1	73.5	50.0
FedProto	79.4	75.8	77.2	60.0
FedDF	81.1	77.6	79.0	63.0
Fed-KD	82.5	79.0	80.8	66.0
FedSlim	81.7	78.2	79.6	64.0
Proposed	89.2	84.6	87.4	83.0

Figure 13 show that acoustic single-modality diagnosis generally performs worse than vibration single-modality diagnosis, indicating that vibration signals contain more direct discriminative information for most mechanical faults, whereas acoustic signals are more susceptible to propagation paths and environmental noise.

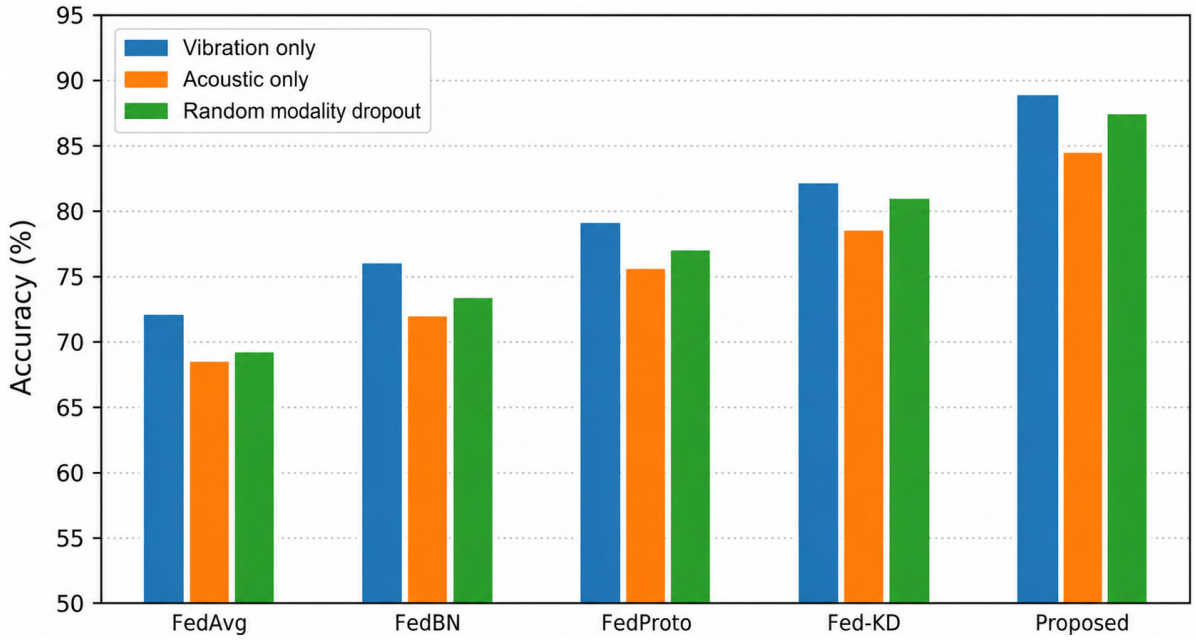

Figure 13. Performance comparison of representative methods under modality-missing conditions at the deployment end.

After model size reduction, FedSlim achieves a slightly lower accuracy than Fed-KD, suggesting that conventional model slimming may lead to the loss of cross-modal knowledge. The proposed method achieves the best results under all three input conditions, with an accuracy of 89.2% for vibration-only input, 84.6% for acoustic-only input, and 87.4% under random modality-missing conditions. These results indicate that input-sensitive relations help the student model learn the different contributions of each modality to fault discrimination, while the gating mechanism preserves critical channels and suppresses redundant ones, thereby improving robustness at the deployment end. This setting also

provides a cross-device-like validation because different operating frequencies are assigned to different clients and the 30 Hz condition is completely unseen during training.

To further clarify the rare-fault recognition capability of the proposed method, the rare-fault results under different challenging scenarios are summarized in Table 10. These scenarios include composite imbalance in the CWRU dataset, modality-fragmented rare-fault diagnosis in the HUST motor dataset, and unseen operating-condition testing on the HUST motor dataset. Fed-KD is selected as the main reference method because it is the strongest distillation-based baseline in most rare-fault settings.

Table 10. Summary of rare-fault validation under different challenging settings.

Dataset	Setting	Rare Fault	Method	Avg Acc/%	Macro-F1/%	Rare-Fault Recall/%
CWRU	Composite imbalance	OR14	Fed-KD	74.7	72.0	58.0
			Proposed	84.9	84.0	80.0
HUST motor	Modality-fragmented rare fault	BRB	Fed-KD	81.2	80.0	66.0
			Proposed	88.4	87.8	84.0
	Unseen 30 Hz condition		Fed-KD	76.8	75.1	61.0
			Proposed	84.3	83.5	78.0

As shown in Table 10, the proposed method consistently outperforms Fed-KD in rare-fault recall under different challenging settings. In the CWRU composite imbalance scenario, the OR14 recall is improved from 58.0% to 80.0%. In the HUST modality-fragmented rare-fault scenario, the BRB recall is improved from 66.0% to 84.0%. Under the unseen 30 Hz operating condition, the BRB recall is improved from 61.0% to 78.0%. These results indicate that the proposed method improves not only average accuracy, but also minority-fault recognition under sample imbalance, modality fragmentation, and unseen-condition generalization.

To verify that the reported improvements are not caused by a specific random initialization or client partition, the proposed method and the strongest baseline were compared over five repeated runs. The results are summarized in Table 11.

Table 11. Statistical comparison over five repeated runs.

Dataset	Scenario	Metric	Fed-KD	Proposed	*p*-Value
CWRU	Composite imbalance	OR14 Recall/%	58 ± 2.6	80 ± 1.8	<0.01
		Macro-F1/%	72.0 ± 1.5	84.0 ± 1.1	<0.01
HUST motor	Modality-fragmented rare fault	BRB Recall/%	66.0 ± 2.3	84.0 ± 1.6	<0.01
HUST motor	Unseen 30 Hz condition	BRB Recall/%	61.0 ± 2.8	78.0 ± 2.0	<0.01

In Table 11, the proposed method consistently outperforms Fed-KD over repeated runs. The improvements in rare-fault recall and Macro-F1 are statistically significant under the main challenging scenarios, indicating that the gains are not caused by random initialization or a single client partition.

4.6. Lightweight and Deployment Performance Analysis

In addition to diagnostic performance, the model complexity and deployment cost of the main federated methods are further compared. Since the HUST motor dataset contains dual-modal inputs, its deployment complexity is higher than that of the single-modal

CWRU experiment. Therefore, this experiment better reflects the practical significance of the dynamic slimming mechanism. The experimental results are shown in Table 12.

Table 12. Model complexity and deployment overhead in the HUST motor experiment.

Methods	Parameter Count/M	FLOPs/M	Time of Diagnosis /ms	Memory Usage /MiB
FedAvg	1.42	3.84	4.26	965.4
FedBN	1.42	3.84	4.18	948.7
FedProto	1.35	3.62	4.03	922.5
FedDF	1.58	4.18	4.61	1018.3
Fed-KD	1.21	3.10	3.72	835.6
FedSlim	0.74	1.82	2.64	628.2
Proposed	0.68	1.55	2.31	596.4

Table 12 further shows that FedDF has the largest number of parameters and the highest GPU memory consumption because server-side ensemble distillation requires maintaining relatively complex models. Fed-KD reduces model complexity to some extent, but its student model still needs to preserve relatively complete feature channels to fit the teacher outputs. FedSlim achieves lower complexity; however, due to the lack of a relational knowledge preservation mechanism, its accuracy decreases under modality-missing conditions. The proposed method outperforms Fed-KD in terms of parameter size, FLOPs, inference time, and GPU memory consumption, and it is slightly better than FedSlim, while maintaining the highest diagnostic performance. These results indicate that the gated student model does not simply remove parameters but selectively preserves critical diagnostic knowledge under the constraint of relational distillation, thereby achieving a balance between performance and efficiency.

4.7. Ablation and Sensitivity Analysis

To isolate the contribution of each module, ablation studies were conducted by removing one component from the complete model at a time. The evaluated components included output-discriminative knowledge, class-prototype relations, input-sensitive relations, rare-fault-aware weighting, and gated slimming. The CWRU composite imbalance scenario was used as the main ablation setting because it contains class absence, low-quality rare samples, and majority-class dominance simultaneously. The results are shown in Table 13.

Table 13. Ablation study of the proposed method on the CWRU composite imbalance scenario.

Variant	Avg Acc/%	Macro-F1/%	OR14 Recall/%	Parameter Ratio
Fed-KD baseline	74.7	72.0	58.0	0.85
W/o output knowledge	82.4	81.0	74.0	0.48
W/o prototype relation	81.8	80.4	72.0	0.48
W/o sensitivity relation	82.1	80.7	71.0	0.48
W/o rare-fault-aware weighting	81.5	80.2	68.0	0.48
W/o gated slimming	85.8	84.9	82.0	1.00
Proposed	84.9	84.0	80.0	0.48

As shown in Table 13, removing any component leads to a decrease in diagnostic performance or deployment efficiency. Removing output knowledge reduces the overall accuracy and Macro-F1, indicating that soft decision responses are important for federated knowledge transfer. Removing prototype relations decreases OR14 recall from 80.0% to 72.0%, showing that class structure preservation is necessary for distinguishing adjacent fault classes. Removing input-sensitive relations also reduces OR14 recall, indicating

that sensitivity responses help preserve weak rare-fault features. The largest drop in rare-fault recall is observed when rare-fault-aware weighting is removed, confirming that minority-fault knowledge can be suppressed during conventional federated fusion. When gated slimming is removed, the model achieves slightly higher accuracy but requires a full parameter structure. Therefore, the gated slimming module improves deployment efficiency while maintaining rare-fault recognition performance.

To further verify the weighting design in Equation (10), a compact sensitivity analysis was conducted on the CWRU composite imbalance scenario, as shown in Table 14.

Table 14. Sensitivity analysis of the rare-fault-aware weighting coefficients.

Setting	$(\beta_r, \beta_\rho, \beta_d)$	Avg Acc/%	Macro-F1/%	OR14 Recall/%
W/o scarcity term	(0, 1.0, 0.5)	83.1	82.2	72.0
W/o reliability term	(1.0, 0, 0.5)	82.7	81.8	70.0
Weak dispersion penalty	(1.0, 1.0, 0.25)	84.3	83.4	78.0
Selected setting	(1.0, 1.0, 0.5)	84.9	84.0	80.0
Strong dispersion penalty	(1.0, 1.0, 1.0)	84.1	83.1	76.0

As shown in Table 14, removing the scarcity or reliability term clearly reduces OR14 recall, indicating that both minority-class enhancement and confidence-based reliability are important for rare-fault preservation. A weak dispersion penalty may allow unstable local knowledge to enter the global model, while a strong dispersion penalty may suppress useful rare-fault knowledge from small-sample clients. The selected setting achieves the best balance among average accuracy, Macro-F1, and rare-fault recall.

5. Conclusions and Future Works

This paper proposes a rare-fault-preserving relation knowledge-based federated dynamic slimming method to address data heterogeneity, weakened minority-class knowledge, model redundancy, and high edge-deployment costs in multi-cliental motor fault diagnosis. The method constructs multi-granularity diagnostic knowledge using output-discriminative knowledge, class-prototype relations, and input-sensitive relations and introduces rare-fault-aware weighting to enhance the contribution of reliable minority-class knowledge during federated fusion. For model lightweighting, a redundancy-gated student model slimming mechanism is designed to retain key diagnostic channels under relation distillation constraints, while a rare-fault recall preservation criterion prevents performance degradation caused by excessive compression. Experiments on the CWRU and HUST motor datasets show that the proposed method outperforms compared methods under composite imbalance, cross-condition generalization, and modality-missing deployment scenarios. Confusion matrices and t-SNE results further verify its effectiveness in reducing rare-fault misclassification and improving feature separability, while complexity analysis demonstrates reduced parameter count, inference time, and GPU memory usage, indicating good potential for edge deployment.

The current validation is limited to a small number of clients and controlled non-IID settings. In ultra-large-scale or extreme non-IID scenarios, relation knowledge fusion may face higher communication costs and reduced reliability due to incomplete fault categories and biased operating conditions. In online federated learning, new faults and data drift may further affect the stability of learned relations. Future work will focus on scalable aggregation, uncertainty-aware weighting, drift-aware updating, and validation on in-house or industrial multi-client motor datasets.

Author Contributions: Methodology, G.Z. and J.Z.; writing—original draft, G.Z.; writing— review and editing, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this article have been presented in the article.

Acknowledgments: It should be noted that Figures 1–4 in the manuscript were drawn with the help of the generative artificial intelligence tool ChatGPT 5.5 to increase the readability and clarity of the algorithm implementation process.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yousuf, M.; Alsuwian, T.; Amin, A.A.; Fareed, S.; Hamza, M. IoT-based health monitoring and fault detection of industrial AC induction motor for efficient predictive maintenance. *Meas. Control* **2024**, *57*, 1146–1160.
2. Evangeline, S.I.; Darwin, S.; Raj, E.F.I. A deep residual neural network model for synchronous motor fault diagnostics. *Appl. Soft Comput.* **2024**, *160*, 111683. [CrossRef]
3. Kim, M.C.; Lee, J.H.; Wang, D.H.; Lee, I.S. Induction motor fault diagnosis using support vector machine, neural networks, and boosting methods. *Sensors* **2023**, *23*, 2585. [CrossRef] [PubMed]
4. Deng, J.; Cheng, Z.; Gu, A.; Zhang, S. Research on equipment fault diagnosis model based on gan and inverse PINN: Solutions for data imbalance and rare faults. *PLoS ONE* **2025**, *20*, e0324180.
5. Chen, H.; Liu, R.; Xie, Z.; Hu, Q.; Dai, J.; Zhai, J. Majorities help minorities: Hierarchical structure guided transfer learning for few-shot fault recognition. *Pattern Recognit.* **2022**, *123*, 108383. [CrossRef]
6. Qiu, S.; Cui, X.; Ping, Z.; Shan, N.; Li, Z.; Bao, X.; Xu, X. Deep learning techniques in intelligent fault diagnosis and prognosis for industrial systems: A review. *Sensors* **2023**, *23*, 1305. [CrossRef]
7. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. *Knowl.-Based Syst.* **2021**, *213*, 106679. [CrossRef]
8. Han, J.; Zhang, X.; Xie, Z.; Zhou, W.; Tan, Z. Federated learning-based equipment fault-detection algorithm. *Electronics* **2024**, *14*, 92.
9. Yao, D.; Liu, H.; Yang, J.; Li, X. A lightweight neural network with strong robustness for bearing fault diagnosis. *Measurement* **2020**, *159*, 107756. [CrossRef]
10. Zhang, J.; Zhao, Z.; Jiao, Y.; Jiao, Y.; Zhao, R.; Xu, X.; Che, R. DPCCNN: A new lightweight fault diagnosis model for small samples and high noise problem. *Neurocomputing* **2025**, *626*, 129526. [CrossRef]
11. Bai, R.; Li, Y.; Qiao, B.; Wang, X.; Wang, T.; Noman, K. FedDwa: A lightweight federated learning with dynamic weighted average aggregation method for machines RUL prediction. *Adv. Eng. Inform.* **2026**, *69*, 104075. [CrossRef]
12. Zhao, C.; Shen, W. A federated distillation domain generalization framework for machinery fault diagnosis with data privacy. *Eng. Appl. Artif. Intell.* **2024**, *130*, 107765.
13. Sun, W.; Yan, R.; Jin, R.; Zhao, R.; Chen, Z. FedAlign: Federated model alignment via data-free knowledge distillation for machine fault diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 3506112.
14. Zhou, Y.; Wang, J.; Wang, Z. Bearing faulty prediction method based on federated transfer learning and knowledge distillation. *Machines* **2022**, *10*, 376. [CrossRef]
15. Zhang, T.; Chen, J.; Li, F.; Zhang, K.; Lv, H.; He, S.; Xu, E. Intelligent fault diagnosis of machines with small & imbalanced data: A state-of-the-art review and possible extensions. *ISA Trans.* **2022**, *119*, 152–171. [PubMed]
16. Dantas, P.V.; Sabino da Silva, W., Jr.; Cordeiro, L.C.; Carvalho, C. A comprehensive review of model compression techniques in machine learning: PV Dantas et al. *Appl. Intell.* **2024**, *54*, 11804–11844.
17. Gong, R.; Wang, C.; Li, J.; Xu, Y. Lightweight fault diagnosis method in embedded system based on knowledge distillation. *J. Mech. Sci. Technol.* **2023**, *37*, 5649–5660. [CrossRef]
18. Sun, J.; Liu, Z.; Wen, J.; Fu, R. Multiple hierarchical compression for deep neural network toward intelligent bearing fault diagnosis. *Eng. Appl. Artif. Intell.* **2022**, *116*, 105498. [CrossRef]
19. Zhang, W.; Biswas, G.; Zhao, Q.; Zhao, H.; Feng, W. Knowledge distilling based model compression and feature learning in fault diagnosis. *Appl. Soft Comput.* **2020**, *88*, 105958. [CrossRef]
20. Ji, M.; Peng, G.; Li, S.; Cheng, F.; Chen, Z.; Li, Z. A neural network compression method based on knowledge-distillation and parameter quantization for the bearing fault diagnosis. *Appl. Soft Comput.* **2022**, *127*, 109331. [CrossRef]
21. Liao, J.; Wei, S.; Xie, C.; Zeng, T.; Sun, J.; Zhang, S.; Zhang, X. BearingPGA-Net: A lightweight and deployable bearing fault diagnosis network via decoupled knowledge distillation and FPGA acceleration. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 3506414. [CrossRef]

22. Hinton, G.; Vinyals, O.; Dean, J. Distilling the knowledge in a neural network. *arXiv* **2015**, arXiv:1503.02531. [CrossRef]
23. Gou, J.; Yu, B.; Maybank, S.J.; Tao, D. Knowledge distillation: A survey. *Int. J. Comput. Vis.* **2021**, *129*, 1789–1819. [CrossRef]
24. Lu, R.; Liu, S.; Gong, Z.; Xu, C.; Ma, Z.; Zhong, Y.; Li, B. Lightweight knowledge distillation-based transfer learning framework for rolling bearing fault diagnosis. *Sensors* **2024**, *24*, 1758. [CrossRef] [PubMed]
25. Neogi, D.; Das, N.; Deb, S. FitNet: A deep neural network driven architecture for real time posture rectification. In *Proceedings of the 2021 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*; NIT: The Woodlands, TX, USA, 2021; pp. 354–359.
26. Park, W.; Kim, D.; Lu, Y.; Cho, M. Relational knowledge distillation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3967–3976.
27. Qiao, Z.; Ning, S.; Gai, Y.; Xie, C. A digital twin guided physical-virtual denoising method for early fault detection of rolling element bearings. *Mech. Syst. Signal Process.* **2026**, *249*, 114108. [CrossRef]
28. Ning, S.; Qiao, Z.; Peng, B.; Zhu, R.; Zhang, C. A digital twin-driven cross-domain adaptation method for bearing intelligent fault diagnosis. *Nondestruct. Test. Eval.* **2026**, 1–23. [CrossRef]
29. Chen, J.; Li, J.; Huang, R.; Yue, K.; Chen, Z.; Li, W. Federated transfer learning for bearing fault diagnosis with discrepancy-based weighted federated averaging. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3514911. [CrossRef]
30. Li, X.; Zhang, C.; Li, X.; Zhang, W. Federated transfer learning in fault diagnosis under data privacy with target self-adaptation. *J. Manuf. Syst.* **2023**, *68*, 523–535. [CrossRef]
31. Zhao, K.; Hu, J.; Shao, H.; Hu, J. Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy. *Reliab. Eng. Syst. Saf.* **2023**, *236*, 109246. [CrossRef]
32. Xue, X.; Zhao, X.; Zhang, Y.; Ma, M.; Bu, C.; Peng, P. Federated transfer learning with consensus knowledge distillation for intelligent fault diagnosis under data privacy preserving. *Meas. Sci. Technol.* **2024**, *35*, 015108. [CrossRef]
33. Smith, W.A.; Randall, R.B. Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study. *Mech. Syst. Signal Process.* **2015**, *64*, 100–131. [CrossRef]
34. Zhao, C.; Shen, W.; Zio, E.; Ma, H. Multimodal unified generalization and translation network for intelligent fault diagnosis under dynamic environments. *Eng. Appl. Artif. Intell.* **2025**, *162*, 112559. [CrossRef]
35. Das, A.; Saha, D. Fedprox-based federated transfer learning for efficient model personalization in healthcare. In *Proceedings of the 2025 International Conference on Ambient Intelligence in Health Care (ICAIHC)*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6.
36. Tan, Y.; Long, G.; Liu, L.; Zhou, T.; Lu, Q.; Jiang, J.; Zhang, C. Fedproto: Federated prototype learning across heterogeneous clients. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 8432–8440. [CrossRef]
37. Gao, M.; Zheng, H.; Lin, J.; Feng, X.; Chen, Y. FedKD: A Fine-Grained Parameter-Efficient Federated Co-Tuning Framework With Knowledge Decoupling for Large and Small Foundation Models. *IEEE Trans. Mob. Comput.* **2026**, 1–17. [CrossRef]
38. Li, D.; Lin, W.; Duan, W.; Liu, B.; Chang, V. EarlyBirdFL: Leveraging Early Bird Ticket Networks for Enhanced Personalized Learning. *IEEE Trans. Emerg. Top. Comput. Intell.* **2024**, *9*, 2879–2893. [CrossRef]
39. Li, X.; Jiang, M.; Zhang, X.; Kamp, M.; Dou, Q. Fedbn: Federated learning on non-iid features via local batch normalization. *arXiv* **2021**, arXiv:2102.07623.
40. Lin, T.; Kong, L.; Stich, S.; Jaggi, M. Ensemble distillation for robust model fusion in federated learning. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 2351–2363.
41. Zhu, Z.; Hong, J.; Drew, S.; Zhou, J. Resilient and communication efficient learning for heterogeneous federated systems. *Proc. Mach. Learn. Res.* **2022**, *162*, 27504. [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.