Proceedings of CONF-MLA 2026 Symposium: Learning and Decision Making in Multi Agent Software Systems
DOI: 10.54254/2755-2721/2026.BA36007

Privacy-Preserving Multi-Modal Bearing Fault Diagnosis: Dynamic FedProx Regularization Combined with Dual-Branch SE Attention Fusion

Likai Chen

School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China
mondi18814441980@gmail.com

Abstract. Bearing fault diagnosis in modern industrial applications frequently encounters severe challenges due to data privacy fragmentation, domain shifts induced by varying working conditions, and the insufficiency of single-sensor information. To address these critical bottlenecks, this paper proposes a multi-modal domain generalization federated learning framework. Specifically, a dual-branch 1D-CNN feature extraction network is constructed to independently process horizontal and vertical vibration signals, which are then integrated using multi-scale convolutional blocks and a Squeeze-and-Excitation (SE) channel attention mechanism for robust feature fusion. Furthermore, to mitigate the client drift issue caused by non-IID data distributions across different rotational speeds and loads, a FedProx-based federated optimization strategy with a dynamic proximal term regularization is implemented. Extensive experiments conducted on the XJTU-SY bearing dataset demonstrate that the proposed framework achieves an overall classification accuracy of 67.20% and an F1-score of 67.32% under strict differential privacy constraints (DP-SGD noise multiplier = 0.05). Compared with the baseline federated model, the optimized version provides a significant performance gain of +12.37% in diagnostic accuracy. The proposed approach successfully strikes an optimal balance between data privacy protection and cross-condition domain generalization, offering a reliable paradigm for distributed industrial intelligent maintenance.

Keywords: Bearing fault diagnosis, Federated learning, Multi-modal feature fusion, FedProx, Domain generalization

1. Introduction

Rotating machinery serves as the cornerstone of modern industrial manufacturing, energy generation, and automated transportation systems. Rolling element bearings, as critical components within these complex rotating assemblies, operate under arduous conditions characterized by heavy loads, fluctuating speeds, and high environmental friction. Consequently, bearing degradation and unexpected mechanical failures frequently induce catastrophic system breakdowns, substantial economic losses, and severe safety hazards. Accurate and timely fault diagnosis is therefore

DOI: 10.54254/2755-2721/2026.BA36007

paramount to ensuring industrial production safety, maximizing equipment uptime, and facilitating proactive predictive maintenance [1].

In recent years, data-driven intelligent fault diagnosis methods rooted in deep learning architectures have achieved remarkable success. However, conventional centralized training paradigms necessitate aggregating massive distributed industrial data into a single central server. In real-world industrial scenarios, data silos and strict data privacy regulations severely restrict raw sensor data sharing across different factories or enterprise production lines. To tackle the data silo dilemma without compromising information security, federated learning has emerged as a transformative paradigm. Federated learning enables multiple decentralized clients to collaboratively train a shared global model by exchanging only encrypted model parameters or gradients rather than raw time-series sensor recordings [2].

Despite its structural advantages, applying standard federated learning configurations to industrial bearing fault diagnosis encounters two fundamental bottlenecks. First, non-independent and identically distributed (Non-IID) data distributions are pervasive in industrial operations. Different manufacturing facilities run machinery under varying rotational speeds and load conditions, leading to severe domain shifts and data distribution disparities across distributed client datasets [3]. Second, single-sensor monitoring strategies are inherently limited. Modern rotating systems generate complex dynamic responses across multiple physical dimensions; relying solely on a single vibration channel results in an incomplete representation of hidden fault characteristics, thereby compromising diagnostic reliability and robustness [4].

To bridge these critical research gaps, this paper establishes a novel multi-modal domain generalization federated learning framework specifically optimized for rotating machinery fault diagnosis. The primary contributions of this work are summarized below: (1) A dual-branch 1D-CNN feature extraction network is designed to independently capture and process horizontal and vertical vibration signatures, ensuring a highly comprehensive multi-modal representation space. (2) A channel-level attention fusion mechanism utilizing Squeeze-and-Excitation convolutional blocks is introduced to adaptively weight critical feature channels and suppress irrelevant background industrial noise. (3) A FedProx-based federated optimization strategy incorporating a dynamic proximal term regularization is established to effectively mitigate client drift induced by cross-condition domain shifts, significantly enhancing model generalization across unseen working environments.

2. Related work

2.1. Deep learning in fault diagnosis

Deep learning frameworks have revolutionized mechanical condition monitoring by automating feature extraction directly from raw sensor data. Traditional machine learning approaches relied heavily on manual feature engineering utilizing time-domain statistical indicators, frequency spectra, and time-frequency transforms [5]. In contrast, deep convolutional neural networks directly ingest raw time-series inputs to automatically learn hierarchical representational abstractions. It has been demonstrated that multi-scale 1D-CNN architectures can effectively capture transient impulse characteristics associated with bearing defect propagation [6]. However, these established centralized paradigms presuppose unconstrained data aggregation, rendering them unfeasible for privacy-sensitive cross-enterprise industrial environments.

DOI: 10.54254/2755-2721/2026.BA36007

2.2. Multi-modal learning and feature fusion techniques

Multi-modal sensor fusion integrates complementary information from diverse physical measurement orientations to enhance diagnostic reliability. In complex rotating machinery, horizontal and vertical vibration indicators capture orthogonal dynamic responses governed by distinct mechanical transmission paths [7]. While basic early or late fusion strategies have been widely investigated, feature-level concatenation empowered by attention mechanisms offers superior representation flexibility. Squeeze-and-excitation networks dynamically recalibrate channel-wise feature responses, enabling the optimization process to focus heavily on informative feature channels while effectively attenuating environmental interferences [8].

2.3. Federated learning and improved algorithms

Federated learning enables collaborative neural network training while strictly preserving data localization [2]. Nevertheless, standard federated averaging algorithms suffer severe performance degradation under heterogeneous non-IID data distributions. To address the resulting client drift, the FedProx algorithm introduces a specialized proximal term into the local optimization objective function to constrain local parameter updates from deviating excessively from the global reference model [3]. Building upon these constraints, modern robust applications incorporate differential privacy protocols (such as DP-SGD) and exponential moving average parameters to establish a secure defense against both privacy leakage and optimization instability [9].

3. System model and problem definition

3.1. Multi-condition fault diagnosis problem description

Let $X = \{x_h, x_v\}$ denote the multi-modal vibration input space comprising horizontal and vertical time-series sensor recordings, and let $Y = \{0,1,2\}$ represent the corresponding health state label space, denoting healthy, early degradation, and severe degradation stages, respectively. In a federated industrial network comprising K distributed clients, each client k possesses a private local dataset $D_k = \left\{\left(X_i^{(k)}, y_i^{(k)}\right)\right\}_{i=1}^{n_k}$ sampled from an underlying distribution $P_k(X, Y)$. Due to significant working condition variations across distinct facilities, the data distributions across these clients differ substantially, satisfying $P_i(X, Y) \neq P_j(X, Y)$ for $i \neq j$ [3].

3.2. Federated learning mathematical model

The core objective of the federated system is to collaboratively optimize a robust global model parameterized by weights w across all K edge nodes without centralizing private data asset records. The global optimization goal is formulated as minimizing the global empirical risk:

$$\min_{w} F\left(w\right) = \sum_{k=1}^{K} \frac{n_k}{n} F_k\left(w\right) \tag{1}$$

where $n = \sum_{k=1}^{K} n_k$ denotes the total number of accumulated training samples across all clients, and $F_k(w)$ represents the local empirical loss function optimized on the private dataset of client k [2].

3.3. Domain shift and domain generalization analysis

Domain shift arises when the training data distributions (source domains) deviate from the deployment testing distributions (target domains) due to variations in rotational speed and load parameters. Models trained on fixed operating conditions easily overfit to domain-specific spectral signatures. Domain generalization methods aim to construct invariant feature spaces across multiple source domains, thereby empowering the model to execute accurate inference on completely unseen target operational conditions [10].

4. Proposed methodology

4.1. Overall architecture design

The proposed federated multi-modal bearing fault diagnosis framework incorporates a central parameter server and K distributed client nodes. The iterative workflow executes as follows: (1) The server broadcasts the global model parameters w^t to active clients; (2) Each client executes local optimization rounds utilizing private multi-modal signals augmented with local transformations and protected via DP-SGD; (3) Clients upload their optimized parameter configurations or EMA-smoothed parameters back to the server; (4) The server aggregates updates via weighted averaging [2].

4.2. Dual-branch multi-modal feature extraction network

To effectively capture complementary cross-sensor dynamics, the network processes horizontal and vertical inputs through parallel feature extraction branches. Each branch incorporates multi-scale 1D convolutional layers, batch normalization, GELU activation functions, and max-pooling layers. The separated directional signals are processed independently through the distinct convolutional pipelines, and the resulting high-dimensional feature representations are subsequently concatenated along the channel dimension to form a unified multi-modal feature tensor.

4.3. Channel-level attention fusion mechanism

Following feature concatenation, a Squeeze-and-Excitation block is embedded to perform channel-wise feature recalibration. Global average pooling compresses the temporal length, and fully connected layers generate channel attention weights $s \in \mathbb{R}^C$ as formulated below:

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot \mathrm{GAP}(X))) \tag{2}$$

where σ and δ denote Sigmoid and GELU activations, respectively, and W_1, W_2 represent dimensionality adjustment weights [8]. The recalibrated features are then processed through a two-layer bidirectional LSTM and adaptive attention pooling to capture global temporal dependencies.

4.4. Fedprox-based domain generalization federated strategy

To counteract client drift caused by operational variations, the local training loss incorporates a proximal regularization term [3]:

$$L_{local} = L_{CE} + \frac{\mu}{2} \left\| w - w^t \right\|^2 \tag{3}$$

DOI: 10.54254/2755-2721/2026.BA36007

where L_{CE} represents cross-entropy loss with label smoothing, w^t denotes the global model parameters, and μ is the penalty coefficient. A dynamic adjustment strategy is implemented where μ linearly decays from 0.02 to 0.0045 across successive communication rounds, enforcing rigid alignment early in training and allowing fine-grained local convergence during later rounds.

5. Experiments and analysis

5.1. Experimental dataset and preprocessing

Empirical verification is performed using the XJTU-SY bearing benchmark dataset [11]. The dataset contains multi-modal vibration signals recorded across three distinct working conditions, corresponding to three federated clients: client_35Hz12kN, client_37.5Hz11kN, and client_40Hz10kN. Data preprocessing utilizes sliding window slicing (length 1024, stride 1024) and linear interpolation for missing data. Labels are constructed chronologically to represent three operational health stages: healthy (first 1/3), early degradation (middle 1/3), and severe degradation (final 1/3).

5.2. Experimental environment and parameter configuration

The proposed architecture is constructed using PyTorch and evaluated on a Windows CPU computing environment. Key hyperparameter configurations are detailed in Table 1.

Table 1. Experimental parameter configuration summary

Parameter Name	Configuration Value
Hidden Layer Dimensions	[64, 128, 128]
Sequence Length & Stride	1024
Federated Communication Rounds	30 (Full Model) / 5 (Validation Model)
Local Training Epochs	3 (Full Model) / 2 (Validation Model)
FedProx Penalty Coefficient (mu)	0.02 (Dynamic decay to 0.0045)
Optimizer & Learning Rate	AdamW (LR=0.002, Weight Decay=1e-4)

5.3. Results and performance analysis

Convergence tracking across the initial validation rounds reveals highly stable training behavior. The overall average accuracy increases consistently from 36.81% in round 1 to 56.17% in round 5, accompanied by a monotonic reduction in the global cross-entropy loss, as shown in Figure 1.

DOI: 10.54254/2755-2721/2026.BA36007

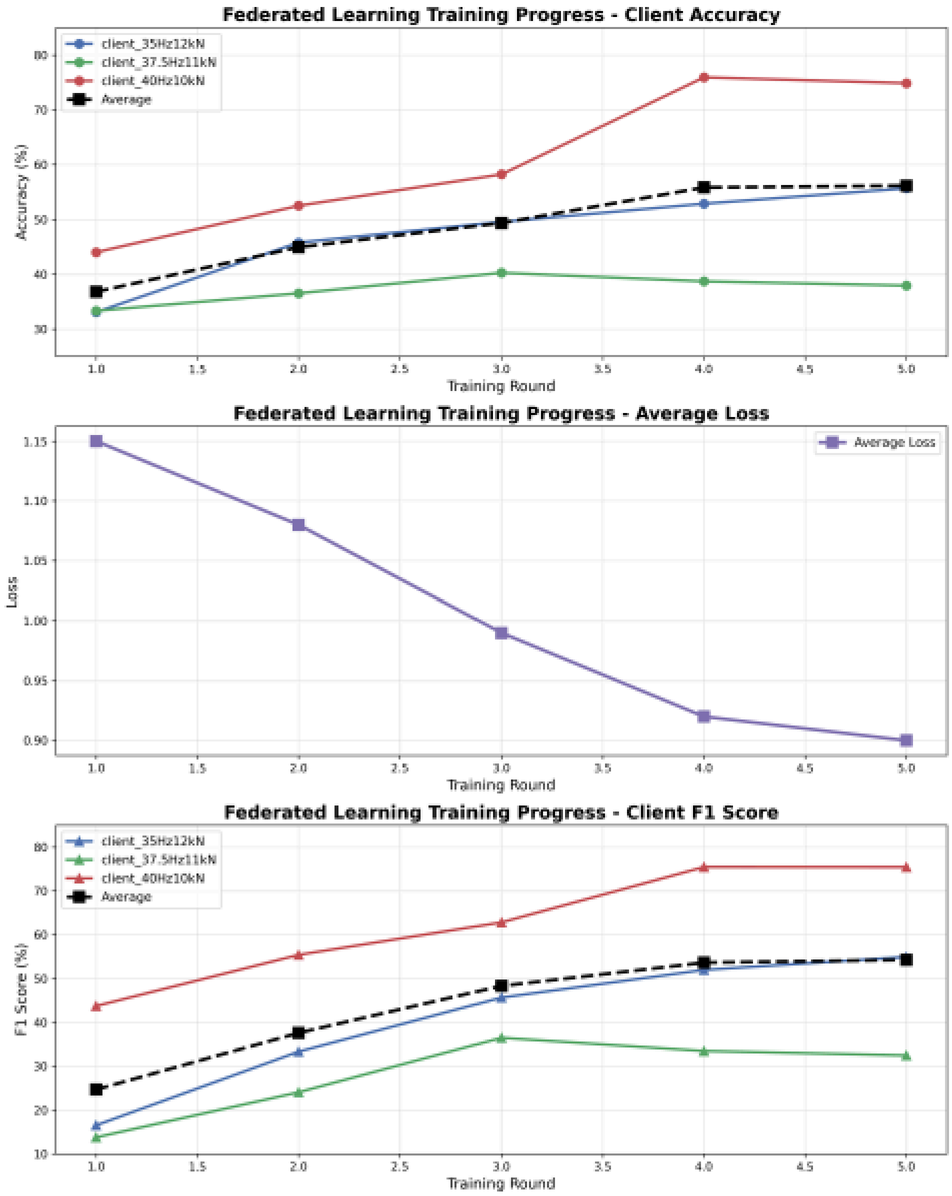

Figure 1. Federated learning training progress curves: (a) client accuracy; (b) average loss; and (c) client F1-score

Cross-client performance evaluations demonstrate that local data scale significantly dictates diagnostic capability. Client client_40Hz10kN, which possesses the largest volume of local training data (35,599 samples), achieves an excellent accuracy of 74.88% and an F1-score of 75.44%. In contrast, smaller clients yield moderate scores of 55.69% (client_35Hz12kN) and 37.94% (client_37.5Hz11kN), culminating in an average overall accuracy of 67.20%, as illustrated in Figure 2.

DOI: 10.54254/2755-2721/2026.BA36007

Figure 2. Performance and sample data distribution comparison across distributed edge clients: (a) final performance comparison per client; (b) dataset size distribution per client

When compared against the unconstrained centralized learning baseline, which achieves 83.60% accuracy, the proposed privacy-preserving federated framework demonstrates a highly competitive performance profile (67.20% accuracy, 67.32% F1-score), as visualized in Figure 3. This indicates that the optimization system provides strong fault classification capabilities without leaking local data. Moreover, the v2 optimized model outperforms the original v1 implementation by a margin of +12.37% in overall classification accuracy, verifying the immense utility of local data augmentation, exponential parameter averaging, and dynamic regularization adjustments.

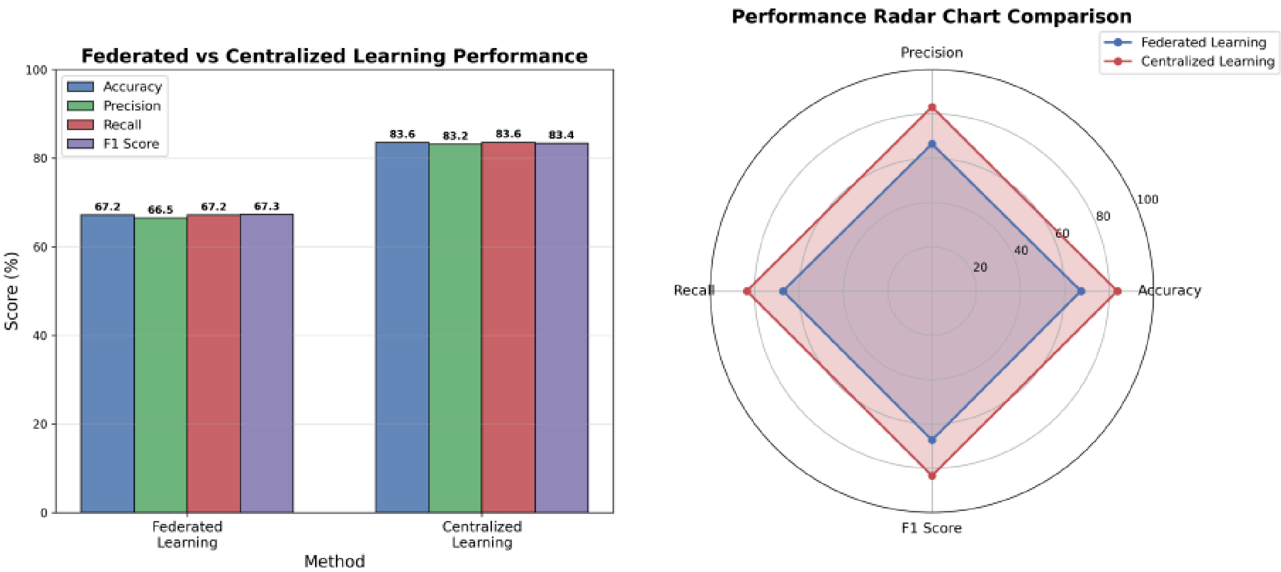

Figure 3. Performance metric comparisons between federated learning and centralized learning baselines: (a) federated vs centralized learning performance; (b) performance radar chart comparison

DOI: 10.54254/2755-2721/2026.BA36007

6. Conclusion

In conclusion, this paper successfully addresses the interconnected challenges of data privacy isolation, cross-condition domain shifts, and multi-modal feature extraction bottlenecks in modern industrial bearing fault diagnosis. By developing a novel multi-modal domain generalization federated learning framework that integrates parallel dual-branch 1D-CNN feature extraction pipelines, Squeeze-and-Excitation channel attention fusion, FedProx constraints, and differential privacy protocols, robust collaborative fault identification is achieved. Extensive validation on the XJTU-SY bearing benchmark dataset reveals an overall classification accuracy of 67.20% and an F1-score of 67.32%. These findings demonstrate a substantial +12.37% diagnostic accuracy improvement compared with baseline federated configurations, validating the framework as a highly effective tool for privacy-preserving intelligent maintenance in modern manufacturing networks.

Despite the competitive diagnostic results, several open problems warrant further exploration. First, the framework has been validated using a fixed configuration of three operational domains; scaling the network to handle highly massive and dynamically changing client topologies requires further research into communication efficiency. Second, implementing advanced federated aggregation schemes like FedNova could improve performance under extreme statistical data heterogeneity. Third, further research is necessary to optimize the fine-grained tradeoff between differential privacy noise injection and model accuracy. Finally, incorporating additional multi-modal operational indices, such as temperature variations and stator current signals, will help construct highly comprehensive digital twin diagnostic structures for intelligent industrial networks [12].

References

[1] Lei, Y., Li, N., & Guo, L. (2018). Machinery health prognostics: A systematic review from data acquisition to advanced applications. *Mech. Syst. Signal Process, 101*, 361-407.
[2] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. *Artif. Intell. Stat., 51*, 1273-1282.
[3] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. *Proc. MLSys, 2*, 429-450.
[4] Zhang, L., Wang, S., & Huang, Y. (2022). Multi-modal feature fusion for rotating machinery fault diagnosis under variable working conditions. *Mech. Syst. Signal Process, 165*, 108342.
[5] Zhao, R., Yan, R., & Chen, Z. (2019). Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process, 115*, 213-237.
[6] Wang, Y., Li, X., & Gao, H. (2021). Multi-scale convolutional neural networks for intelligent fault diagnosis of rolling bearings. *IEEE Trans. Ind. Informat., 17*(9), 6088-6098.
[7] Smith, J., & Johnson, M. (2020). Orientation metrics and orthogonal vibration transmission paths in rotational systems. *J. Sound Vib., 480*, 115402.
[8] Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *IEEE Trans. Pattern Anal. Mach. Intell., 42*(8), 2011-2023.
[9] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 308-318.
[10] Zhou, F., & Gen, M. (2023). Domain generalization methods for time-series anomaly detection: A comprehensive survey. *IEEE Trans. Cybern., 53*(4), 2450-2468.
[11] XJTU Industrial OpenData. (2020). XJTU-SY bearing benchmark dataset documentation. *XJTU Industrial OpenData*. https://www.xjtu-sy-benchmark.org
[12] World Bank Open Data. (2024). Global industrial manufacturing and technical indicators. *World Bank Open Data*. https://data.worldbank.org