Computers and Artificial Intelligence；Vol.3, No.2; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n2.5964
Published by: Zeus Press

Research Review on Federated Learning Technology for Fault Diagnosis

Xuebin Han*

School of Information Science and Engineering, Henan University of Technology, 450001, Zhengzhou, Henan, China

**Corresponding author: Xuebin Han.*

Abstract

Federated learning (FL) is a distributed machine learning (ML) method. This technology only needs to exchange model parameters without sharing private data and plays an important role in industrial fault diagnosis. This paper focuses on analysing four mainstream technical schemes: fault diagnosis methods based on traditional federated learning, fault diagnosis methods based on federated deep learning, optimized federated fault diagnosis methods for nonindependent and identically distributed (non-IID) data, and lightweight federated fault diagnosis methods for edge device deployment. This paper systematically sorts out the federated learning technologies for fault diagnosis, summarizes the practical challenges existing in these technical schemes, and proposes corresponding future research prospects, providing specific references for the innovation and implementation of federated learning technologies for fault diagnosis.

Keywords

federated learning, fault diagnosis, non-IID, privacy protection, lightweight model

1. Introduction

The core task of fault diagnosis in practical industry is to analyse various types of data generated by equipment during operation, such as vibration signals, temperature signals, and current signals, to analyse and monitor the real-time operating status of equipment, thereby identifying relevant fault information, clarifying the fault type and severity, and providing guarantees for operation and maintenance. There are two types of commonly used ML models in fault diagnosis: traditional machine learning models and deep learning models. Traditional machine learning models such as support vector machines (SVMs), random forests, and decision trees have fast training speeds and low computing power requirements, whereas convolutional neural networks (CNNs) and long short-term memory (LSTM) networks have certain computing power requirements, but their diagnostic accuracy is higher.

Federated learning is a distributed machine learning framework [1]. Federated learning can make full use of more data resources to improve model performance without leaking raw data. Traditional machine learning usually requires centralizing data on a central server for training, whereas federated learning does the opposite—keeping data “immobile” and letting models “move”. The general process of federated learning is as follows: In the initial stage of the model, the central server issues an initial model to all participants, allowing each participant to train the model using their own data and calculate parameter updates; participants only upload encrypted model parameter updates to the server instead of raw data; after the calculation is completed, the server uses algorithms such as federated averaging to aggregate the received multiple parameter updates to generate a global update; finally, this global model is issued back to each participant for the next round of training. In this process, through repeated cycles, federated learning can effectively solve the problem of “data islands” [2].

In recent years, research on federated learning technology for fault diagnosis has experienced explosive growth, but existing studies still have many deficiencies. First, the technical schemes are scattered and lack a systematic sorting of the adaptability of technologies in different scenarios; second, the optimization of core pain points such as non-IID data [3] and edge device deployment is still not in-depth enough, and there is no perfect solution; third, most existing review papers list only relevant research results and lack targeted analysis of technical challenges and future development directions. Therefore, this paper starts with the mainstream technical schemes of federated learning for fault diagnosis, compares different technical schemes, and presents development prospects for related difficulties and challenges, hoping to support the innovative breakthroughs of subsequent research.

2. Mainstream Technical Schemes of Federated Learning for Fault Diagnosis

2.1 Fault Diagnosis Methods Based on Traditional Federated Learning

The core idea of fault diagnosis methods based on traditional federated learning is to combine traditional machine learning models such as SVM and logistic regression with the FL framework. This kind of fault diagnosis method based on traditional federated learning retains the advantage of low computing power requirements and maintains a relatively fast training speed to a certain extent. Of course, the diagnostic accuracy of this method is lower than that of federated deep learning methods. Reference [4] proposed a small-sample bearing fault diagnosis method based on short-time Fourier transform and SVM and other technologies. Using the bearing fault dataset (which is from Case Western Reserve University, USA), they reported that in 10 tests of the proposed model, even if the training set of each fault type accounts for only 10% of the total sample proportion, the average accuracy can still reach approximately 92% [4]. Compared with other diagnostic models, this method has higher diagnostic accuracy and stronger feature extraction ability under small-sample conditions.

2.2 Fault Diagnosis Methods Based on Federated Deep Learning

Fault diagnosis methods based on federated deep learning are currently the mainstream research direction in this field. It combines deep learning models such as CNNs and long short-term memory (LSTM) with a federated learning framework, which not only combines the distributed training characteristics of federated learning but can also effectively use deep learning models to extract deep features of data. Therefore, in complex scenarios, we can still maintain a good detection rate and accuracy in fault diagnosis. In reference [4], reference [5] proposed a federated learning method that combines a CNN and Gaussian mixture model and uses the federated averaging (FedAvg) algorithm. Experiments have shown that the accuracy of the FedAvg method on the CWRU dataset reaches 96.10%, the specificity and recall rate are both above 94%, and the comprehensive evaluation index is 94.80% [5]. This clearly proves the robustness of the method.

2.3 Optimized Federated Fault Diagnosis Methods for Non-IID Data

In actual industrial scenarios, fault information is often scarce; that is, it has the characteristics of small samples, and the data distributions of different types of equipment differ. These two reasons increase the training difficulty of federated machine learning models and correspondingly reduce the accuracy of fault diagnosis. Therefore, in terms of the optimization of federated fault diagnosis, we can adopt two strategies: one is to adopt algorithms such as federated transfer learning and dynamic clustering to accelerate the convergence speed of the model; the other is to perform noise addition or reconstruction on the generated fault data for expansion under the federated learning framework and alleviate the small-sample problem by increasing the sample size. Reference [6] adopted the first strategy and proposed an improved FedProx algorithm for the problem of poor model classification performance. The model trained by this improved FedProx algorithm has an mAP0.5 index of 0.9392 [6] on the Non-IID dataset with only a single category for each client, which is 0.0785 and 0.1289 higher than those of the FedAvg and FedProx algorithms, respectively, fully proving the effectiveness of the improved FedProx algorithm.

2.4 Lightweight Federated Fault Diagnosis Methods for Edge Device Deployment

Edge devices such as small sensors in factories have limited computing power and small storage capacity and are obviously not suitable for large-scale iteration of federated models. Reference [7] proposed a lightweight load decomposition framework and method based on federated learning and knowledge distillation. They designed a hybrid architecture that combines a CNN and a transformer. This architecture extracts local time series features through the CNN module, enhances the modelling ability of long-term time series dependencies by using the improved transformer structure, and constructs a cloud-edge collaborative training mode based on federated learning and knowledge distillation. The introduction of a lightweight model as the global model significantly reduces the communication overhead, which is approximately 85% lower than that of traditional federated learning [7], providing an effective solution for deployment in edge computing scenarios. To intuitively compare the core characteristics of the four technical schemes, the key indicators are summarized in Table 1.

Table 1: Comparison of Four Technical Schemes

Type of Technical Scheme	Core Model/Algorithm	Core Advantages	Application Scenarios	Typical Performance	Limitations
Traditional Federated Learning Fault Diagnosis	SVM/Random Forest + Federated Averaging	Low computing power, fast training, small-sample friendly	Simple faults, low-computing edge devices	Bearing fault diagnosis accuracy ≈ 92%	No deep feature extraction ability, low accuracy
Federated Deep Learning Fault Diagnosis	CNN/LSTM + FedAvg/Gaussian Mixture Model	High accuracy, strong robustness, automatic feature extraction	Complex faults, multiworking condition scenarios	CWRU dataset accuracy ≈ 96.10%	High computing/communication overhead, difficult to deploy on small edge devices
Federated Optimized Fault Diagnosis for Non-IID Data	Improved FedProx/Federated Transfer Learning + Data Augmentation	Alleviate data heterogeneity, improve convergence speed	Scenarios with scarce samples and uneven data distribution	mAP0.5 of Non-IID dataset ≈ 0.9392	Some algorithms increase model complexity
Lightweight Federated Fault Diagnosis for Edge Devices	CNN-Transformer + Knowledge Distillation + Federated Learning	Lightweight, low communication overhead	Industrial sensors/small controllers	Communication volume reduced by ≈ 85% compared with traditional FL	May cause a slight decrease in diagnostic accuracy

3. Existing Challenges and Problems

3.1 Data Level

In the current industrial scenarios of federated learning, the dynamic adaptation of non-IID data is obviously insufficient. In the actual production process, the state of equipment is constantly changing and real-time. Changes in working conditions obviously lead to real-time changes in data distribution. In contrast, most existing optimization strategies are designed for static non-IID data, which cannot meet the adaptive requirements in actual demand [8]. In addition, the limitations of small-sample scenarios exist. Compared with other types of data, industrial fault data can be called scarce. This is because the fault conditions are obviously far less than the normal working conditions, which is determined by the sample size. Therefore, the generalization ability of the model in small-sample scenarios still needs to be improved.

3.2 Model Level

As mentioned earlier, federated learning is a distributed framework, which leads to heterogeneous model aggregation. Different types and structures of ML models may be adopted for different engineering participants; that is, the use of models is uncertain and inconsistent. Although breakthroughs have been made in heterogeneous model aggregation, traditional federated learning aggregation algorithms—such as SCAFFOLD and FedProx—all require model isomorphism. In addition, the problem of performance balance cannot be ignored. High-precision privacy protection mechanisms often significantly increase system communication overhead; ensuring the strength of privacy protection without reducing model performance and diagnostic speed is still difficult.

3.3 Engineering Level

The practical challenges at the engineering level include two main aspects: high communication overhead and difficult deployment of edge devices. In terms of communication overhead, federated learning models involve multiple participants, and data transmission among these participants generates high communication costs; the number of parameters and iterations of deep learning models are relatively large, which further puts pressure on communication. In terms of edge device deployment, since federated learning technology has not yet formed a unified technical standard, equipment from different manufacturers often has poor compatibility. Although some progress has been made in lightweight technologies for edge-side target recognition [9], the deployment of edge devices is still difficult. In addition, the centralized architecture of federated learning is prone to single-point failures, which further increase the risk of engineering deployment.

3.4 Security Level

The parameters of federated learning are shared. This multiparty communication mechanism makes it easy for malicious participants to mix in. In the process of multiparty communication, malicious participants may perform poisoning attacks on the model [10] or modify the uploaded model gradients. As a result, it is difficult for the federated server to verify the authenticity of the parameters, the protection cost increases accordingly, and the model performance is also reduced. Undoubtedly, this problem cannot be ignored, but there is still no relatively perfect solution. This is because most existing protection methods are aimed at a single type of attack, and an effective security mechanism that can uniformly protect against multiple types of attacks is lacking.

4. Future Research Prospects

In response to the challenges existing in current research, future research on federated learning technology should focus on four directions: data adaptation, model innovation, technology integration and engineering implementation.

In terms of data adaptation, we should focus on promoting the research of dynamic non-IID data adaptive algorithms and corresponding improved algorithms such as FedRag [11] to better monitor data in real time and simultaneously strengthen small-sample data augmentation technology, combine it with federated meta-learning and contrastive learning, and explore federated-level data quality evaluation mechanisms.

In terms of model innovation, it is necessary to study efficient heterogeneous model [12] aggregation mechanisms to reduce aggregation bias and simultaneously study the collaborative optimization of lightweight and privacy protection [13], study low-overhead privacy protection algorithms, and combine technologies such as knowledge distillation [14] to continuously improve the interpretability of federated models.

In terms of technology integration, the integration of federated learning with emerging technologies such as digital twins and blockchains should be promoted [15], virtual fault data should be generated through digital twins to alleviate small-sample and non-IID problems, an edge-cloud collaborative training framework using edge intelligence should be constructed, and resistance to malicious attacks should be improved, and incentive mechanisms based on blockchain decentralized training should be improved.

In terms of engineering implementation, standardized schemes for federated learning for fault diagnosis should be formulated, low-cost deployment tools should be researched and lightweight model technology should be optimized to reduce the application cost of small and medium-sized enterprises, landing verification of typical equipment such as bearings and motors should be carried out, and large-scale application of the technology should be realized.

5. Conclusion

This paper systematically reviews the research status of FL technology for fault diagnosis, which can provide references for the innovation and engineering implementation of federated learning technology for fault diagnosis. This paper first introduces the core logic of fault diagnosis and the core theory of federated learning; then, four mainstream technical schemes of federated learning for fault diagnosis are introduced, and practical problems and challenges such as insufficient dynamic non-IID data adaptation, difficult heterogeneous model aggregation, difficult edge device deployment, and prominent security risks at the data level, model level, security level and engineering level are listed. In the future, the practical space of this technology can be promoted through data adaptation optimization, model innovation, and the formulation of unified engineering deployment standards, driving the intelligent transformation of the manufacturing industry.

References

[1] Wang, X., Huang, Q. C., & Sun, L. Y. (2025). A Review of Personalized Federated Learning. Journal of Chinese Computer Systems, 1-13. https://doi.org/10.20009/j.cnki.21-1106/TP.2025-0399.

[2] Zhang, S. Q., Wang, H. F., Wang, Z. P., Zhao, P., Liu, Y. H., Chi, Z. H., & Zhao, X. S. (2025). A Review of Federated Learning-Driven Network Intrusion Detection. Computer Engineering and Applications, 1-19.

[3] Feng, W. J., Yang, Y. Q., Jia, Z. H., & Zhang, L. L. (2025). Optimization Methods of Federated Learning Under Non-Independent and Identically Distributed and Long-Tailed Distributions. Journal of Xinjiang University (Natural Science Edition in Chinese and English), 42(04), 425-433. https://doi.org/10.13568/j.cnki.651094.651316.2024.12.30.0001.

[4] Fu, P. F., Sun, A., Yao, K. M., Li, H. J., & Ma, J. L. J. (2025). A Small-Sample Bearing Fault Diagnosis Method Based on STFT-SWT-2DCNN-SVM. In Proceedings of the 36th Chinese Process Control Conference (Volume 2) (pp. 1028-1035). School of Automation and Information Engineering, Sichuan University of Science & Engineering; Sichuan Provincial Key Laboratory of Artificial Intelligence.

[5] Li, H. (2025). Research on Bearing Fault Diagnosis Based on Federated Learning (Master's Thesis). Shijiazhuang Tiedao University. https://doi.org/10.27334/d.cnki.gstdy.2025.000599.

[6] Wang, R. L., Yan, K., Ning, Z. J., & Xiao, X. (2025). Federated Learning-Based Industrial Instrument Localization and Classification Algorithm for Non-IID Data. Instrument Technique and Sensor, (11), 19-24.

[7] Wang, S. X., Cao, Z., Zhao, Q. Y., Feng, X. C., & Rong, C. Y. (2026). A Lightweight Load Decomposition Method Based on Federated Learning and Knowledge Distillation. Journal of Tianjin University (Science and Technology), 59(01), 52-64.

[8] Sun, H. B., Wang, G. C., Zhang, L., Wang, Y., & Guo, Y. A. (2025). Adaptive Selection Strategy of Federated Learning Clients Based on Deep Reinforcement Learning. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 1-13.

[9] Fu, H. Y. (2024). Research on Lightweight Federated Learning Technology for Edge-Side Object Recognition (Master's Thesis). Beijing University of Posts and Telecommunications. https://doi.org/10.26969/d.cnki.gbydu.2024.001699.

[10] Cao, Y. F., Luo, Q. R., & Zhang, L. (2025). Data Mixing Poisoning Attack Based on Federated Learning. Computer Engineering, 1-8. https://doi.org/10.19678/j.issn.1000-3428.0252414.

[11] Shi, X. J., Zhu, X. L., & Xiao, W. T. (2026). FedReg*: Addressing the Challenges in Non-Independent and Identically Distributed Scenarios for Federated Learning (in English). Journal of Donghua University (English Edition), 43(01), 41-49. https://doi.org/10.19884/j.1672-5220.202412011.

[12] Hou, J. K., Guo, R., Zhang, Y. H., & Liu, G. J. (2025). A Personalized Federated Learning Scheme with Differential Privacy for Heterogeneous Data. Application Research of Computers, 1-13. https://doi.org/10.19734/j.issn.1001-3695.2025.09.0411.

[13] Yu, C. Y., Lian, G. Q., & Su, W. (2026). A Review of Key Technologies of Federated Learning in Non-Independent and Identically Distributed Ultra-High Dimensional Scenarios. Information Technology and Standardization, (Z1), 50-56+62.

[14] Wang, J., Peng, Y. J., Zhang, Z. J., Tian, J., Zhu, J. M., & Li, H. T. (2025). A Personalized Differential Privacy Federated Learning Method Based on Knowledge Distillation. Journal of Xidian University, 1-18. https://doi.org/10.19665/j.issn1001-2400.20260108.

[15] Guo, J. L., Chen, Y. Q., Lu, R., & Li, G. C. (2025). A Privacy-Preserving Decentralized Recommendation System Based on Blockchain and Federated Learning. Application Research of Computers, 1-10. https://doi.org/10.19734/j.issn.1001-3695.2025.09.0301.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights