Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication

Shixiang Lu, *Student Member, IEEE,* Zhiwei Gao, *Senior Member, IEEE,* Qifa Xu,
Cuixia Jiang, Aihua Zhang, Xiangxiang Wang

***Abstract*—Privacy protection as a major concern of the industrial big data enabling entities, makes the massive safety-critical operation data of wind turbine, unable to exert its great value because of the threat of privacy leakage. How to improve the diagnostic accuracy of decentralized machines without data transfer remains an open issue, especially these machines are almost accompanied by skewed class distribution in the real industries. In this study, a class-imbalance privacy-preserving federated learning framework for fault diagnosis of decentralized wind turbine is proposed. Specifically, a biometric authentication technique is first employed to ensure that only legitimate entities can access private data and defend against malicious attacks. Then, the federated learning with two privacy-enhancing techniques enables high potential privacy and security in low-trust systems. After that, a solely gradient based self-monitor scheme is integrated to acknowledge the global imbalance information for class-imbalanced fault diagnosis. We leverage a real-world industrial wind turbine dataset to verify the effectiveness of the proposed framework. By comparison with five state-of-the-art approaches and two non-parametric tests, the superiority of the proposed framework in imbalanced classification is ascertained. An ablation study indicates the proposed framework can maintain high diagnostic performance while enhancing privacy protection.**

***Index Terms*—Privacy preserving, class-imbalance classification, federated learning, wind turbine, fault diagnosis.**

I. Introduction

INCREASINGLY stringent privacy protection legislations and increasing data privacy concerns have brought unprecedented challenges to conventional centralized training models [1]. For fear of data leakage, data generating entities (clients) are unwilling to transfer and share real-time data, resulting in the mode of centralized data collection, storage, and processing unsustainable [2]. In the industrial field, as the core asset of an enterprise, once the fault data of machinery is illegal acquisition by competitors, the information of scheduling and production capacity will be exposed, which may cause unpredictable economic costs. To this end, a stringent data protection regulation, that is, no private data is allowed to leave local storage, is emerging in real indusries [3].

Biometric authentication, as a security technique to ensure that only legitimate entities can access private data and defend against malicious attacks, has been widely used in practice, such as autonomous vehicles [4], smart devices [5], and decentralized manufacturing industry [6]. Well-accepted biometric authentication techniques include face recognition authentication [5], iris recognition authentication, and palmprint authentication [7], among which face recognition authentication moves away from sensor-based biometric authentication, making it available for different purposes that previously required specific sensors. Under the biometric authentication gateway, innovative methodologies to reconcile data enabling and privacy security has become an urgent need for academia and industry [8].

Intelligent diagnosis of wind turbine promotes the transformation from traditional industry towards smart manufacturing [9], [10], Recently, the rapid evolution of data-driven methods in fault diagnosis have generated encouraging performance, showcasing data-driven auxiliary systems can assist the operators to achieve near-zero unplanned downtime and predictive maintenance [11]–[13]. However, the decentralized data caused by privacy preservation makes it difficult to realize the method of centralized training with free data access. Thus, as a nascent area of fault diagnosis, the concept of privacy-preserving intelligent diagnosis is proposed to bridge the gap between data empowerment and data protection. This motivates us to realize efficient fault diagnosis of decentralized industrial machinery while protecting privacy.

Federated learning (FL) grabs useful knowledge from decentralized data through collaborative model learning, enhancing data security, privacy and confidentiality [8], [14]. This paradigm is beneficial to centralized fault diagnosis shift towards privacy-preserving intelligent diagnosis [3]. [15] propose a collaborative training framework based on convolution neural network (CNN) and FL for bearing fault diagnosis. Through two mechanical diagnosis cases, [16] validate that the FL with dynamic validation methods can break the isolated data island problem. [3] propose a federated transfer learning method to address the problem of non-independent and identi-

The work is partially supported by the China Scholarship Council, the National Nature Science Foundation of China (61673074 and 72171070), and the Key Research and Development Program of Anhui Province (202004a05020020), and the Alexander von Humboldt Foundation (GRO/1117303 STP) (Corresponding author: Zhiwei Gao.)

S. Lu, Q. Xu, C. Jiang, and X. Wang are with the School of Management, Hefei University of Technology, Hefei, 230009, China. S. Lu also is with Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, UK, and X. Wang also is with Ronds Science & Technology Incorporated Company, Hefei, 230088, China. (e-mail: lushixiang@mail.hfut.edu.cn; xuqifa@hfut.edu.cn; jiangcuixia@hfut.edu.cn; xiagxiang.wang@ronds.com.cn)

Z. Gao is with Faculty of Engineering and Environment, Northumbria University, Newcastle upon Tyne, UK. (e-mail: zhiwei.gao@northumbria.ac.uk)

A. Zhang is with College of Physical Science and Technology, Bohai University, Jinzhou, 121000, China. (e-mail: zhangaihua@qymail.bhu.edu.cn)

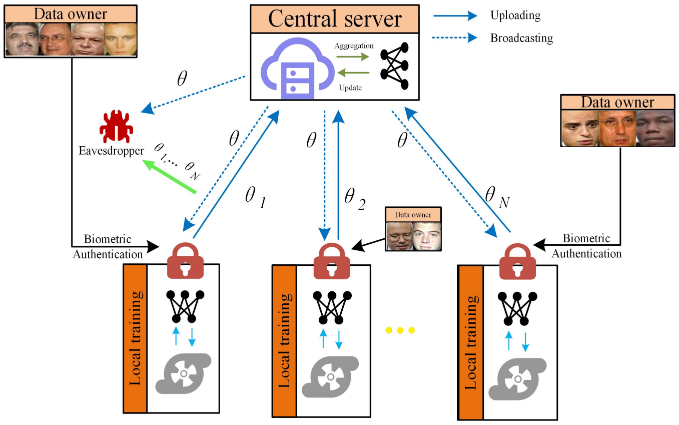

Fig. 1: The schematic of data information inferred by eavesdropping on updated parameters of FL.

cally distributed decentralized fault data. However, FL still has various privacy issues. For example, [17] manipulate a model-inversion attack that reverse engineering updated parameters to reconstruct the instance. As shown in Fig. 1, although each client data can only be accessed by biometric authentication entities, eavesdroppers can still infer the distribution of privacy data from the updated gradient. To this end, [18] design a cryptography computing based FL scheme for fault diagnosis in Internet of Ships. To prevent local updates from being traced, Anonymous uploading and privacy differential strategies are integrated into FL for industrial big data analysis [11]. Although existing FL methods are effective in privacy preserving fault diagnosis, few research results are deployed on wind turbines, because the global balanced data distribution they assumed are far from the real scenarios [9].

In actual industrial applications, wind turbine works in a normal state throughout the entire operating cycle, and failures rarely occur during the operating phase. A research survey provided by Svenska Kullagerfabriken (SKF), a global industrial bearing leader, makes this issue more specific. It is reported that in the whole life cycle of a machine, the ratio of bearing fault/normal is approximately 1/180, which is a typical class imbalanced problem. Under this situation, most existing FL methods may skew to the normal class which accounts for a large proportion of the global training, while the fault class which is more worthy of attention is likely to be ignored because of its small proportion [9], [19]. Addressing the issue of class imbalance can facilitate intelligent diagnostic approaches to enable industrial applications. To this end, cost-sensitive learning is adopted to diagnose imbalance faults of wind turbines [20]. In [21], the authors propose a class rebalancing method to detect blade icing. Additionally, the local training network in each client may encounter over-fitting problem caused by serious class imbalance data [22], thus poisoning the global network and deteriorating the diagnosis accuracy.

To address the issue of class imbalanced learning, a self-balancing FL framework is proposed by rebalancing strategy to relieve the imbalance of training data [2]. However, it increases the computational burden and delays the update of local training. [23] introduce a balanced cross entropy loss function into the FL framework, which only uses shared parameters but with laborious hyperparameter tuning. In contrast, [19] use gradient information to infer global data distribution, which is a promising method. However, the ground-truth gradient of exposure may be tracked by eavesdroppers to infer local client information. In addition, when a client with a small and imbalanced dataset, the problem of gradient explosion may occur. Enlightened by the good generalization of gradient noise mechanism [24], we apply it to address the class-imbalance issue in FL. Another implicit effect is that it provides strict gradient preservation.

In this paper, we propose a novel class-imbalance privacy-preserving federated learning framework, entitled CI-PPFL, for fault diagnosis of decentralized wind turbine. For this purpose, FL firstly learns the decentralized diagnostic knowledge without sharing the private data. Then, a strategy for proportionally updating shared parameters and a gradient noise mechanism are incorporated to prevent tracking the gradient and inferring the data distribution of the clients. Additionally, a gradient self-monitor algorithm detects the imbalanced classes, which dynamically reweights the loss via updating the shared parameters. The effectiveness of CI-PPFL method is investigated on an industrial wind turbine dataset. Our main contributions are summarized as follows.

- A biometric authentication based federated learning framework, namely CI-PPFL, is proposed to address the privacy-preserving issues for fault diagnosis of wind turbine. FL reduces privacy leakage caused by data sharing, and parameter proportional update strategy and gradient noise mechanism hinder gradient tracking and information inference.
- A gradient self-monitor scheme for class-imbalanced fault diagnosis is introduced as well. It acknowledges the global imbalance and updates classification loss by the aggregated gradient solely, without posing threats to privacy.
- A wind turbine imbalanced dataset proves that the proposed CI-PPFL method enables decentralized learning on class-imbalance fault diagnosis problems. The experimental results show that the CI-PPFL can not only enhance the privacy protection, but also significantly improve the performance of dispersive system in dealing with class-imbalance.

The remainder of this article proceeds as follows. In Section II, we briefly review the related work about FL and class-imbalance learning. Section III presents the proposed framework for decentralized fault diagnosis. Section IV conducts the state-of-the-art methods comparison, ablation study, sensitivity analysis and security analysis on an industrial dataset. Section V summarizes and concludes the article.

II. Related work

A. Privacy-preserving federated learning

With the ever-increasing concerns about data security, privacy protection has become a hot and significant issue. FL distributing data-driven models to the clients for decentralized training, rather than centrally aggregating data, is regarded as a promising approach for privacy confidentiality. In recent years, FL has been widely used in data-sensitive real-world scenarios, such as the next word prediction tasks on smartphone [25], segmentation and classification of medical imaging [8], and local governance in the setting of the COVID-19 pandemic [26]. However, recent studies demonstrate that the classical FL is not a completely privacy-preserving approach [27]. [17] perform reverse attack engineering to reconstruct images from shared parameters. To prevent shared parameters from being tracked and used to infer the data distribution of the clients, additional privacy-enhancing techniques should be integrated into FL.

The techniques for privacy-enhancing can be grouped into three categories: anonymization, secure aggregation, and differential privacy (DP). First, anonymization or pseudonymization is utilized to mask the critical data before they are transmitted to the analysis site. For example, [11] anonymize the client participating in the update, so that the eavesdropper cannot map the shared parameters to the client, thus successfully preventing the inference of the client data. Second, secure aggregation realizes parameter transmission in encrypted state by homomorphic encryption [1]. [8] propose a secure aggregation FL framework for non-trivial clinical tasks, and its security is verified in a gradient-based model inversion attack. Third, DP provides quantifiable privacy guarantees by adding a carefully calibrated noise to the gradients [28]. After performing several tasks of medical image segmentation, [27] conclude that the FL with strict privacy guarantee can still achieve excellent classification performance.

B. Class-Imbalance learning

The main challenge of decentralized learning is that the training data on each client is class-imbalance, which results in a deterioration in learning accuracy. The authors are mainly concerned with global imbalance in decentralized learning, that is, the data collection across all clients is class imbalanced. It is quite common in fault diagnosis because faults are rare for every machine, and in most cases a machinery works normally [9]. To tackle the challenge, several progresses have been made in class-imbalance learning. We summarize relevant studies into three categories: data-level, algorithm-level, decision-making-level methods. Data-level method manipulates sample distribution, such as manually rebalancing the training by over-sampling minority classes or under-sampling majority ones [22], [29]. [2] propose a self-balancing FL framework for mobile systems, in which a rebalancing technique is adopted for each client before training begins. In addition to computational burden, this kind of method may also encounter the mismatch between local imbalanced and global imbalanced [19]. Algorithm-level method modifies the learning procedure to make the classifier sensitive toward minority classes, including meta learning [30] and cost-sensitive learning [9], [29], [31]. Among them, re-weighting misclassification loss via inverse class frequency [31] and modifying the loss function are two promising ways. However, aggregating the class frequency of each client may pose threats to privacy. Such as, in [23], a balanced cross entropy loss is integrated into the FL framework to solve the class imbalance problem while reducing privacy leakage. Decision-making-level method adjusts the discriminant probability and tries to move the output threshold toward minority classes [29]. However, defining an optimal cut-point value in multi-classification tasks requires much prior knowledge, which is difficult for decentralized learning.

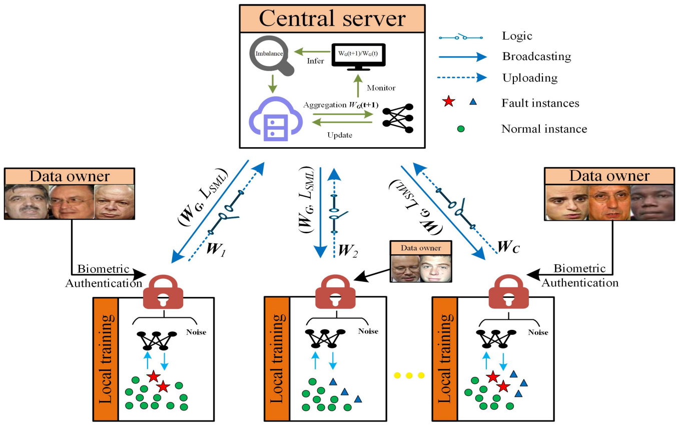

Fig. 2: The proposed CI-PPFL framework.

III. Methodology

In this section, we present the CI-PPFL framework for fault diagnosis. Its principle is shown in Fig. 2. Specifically, biometric authorization technique is adopted to allow only a legitimate server to access local clients. Then, before being sent to logic unit, the locally well-trained parameters will add an appropriate noise mask to prevent the leakage of ground-truth parameters. Further, the logic unit randomly uploads the processed parameters to central server, disturbing the parameter tracking. After that, the uploaded parameters are aggregated to update the global model, and a monitor is used to discover the imbalanced knowledge for modifying the loss function. Finally, the updated parameters and loss function are broadcast to each biometric authentication client for the next iteration.

A. Problem definition

We focus on decentralized fault diagnosis of wind turbine where the classes of instances are severe imbalanced. Let the decentralized diagnosis system S consists of C biometric authentication clients. Any client j contains the same feature space $\boldsymbol{X}$ and label space $\boldsymbol{y} = [1, \cdots, Q]$, where $\boldsymbol{y}$ is a skewed distribution. If we feed the i-th instance $\boldsymbol{X}_i$ of the class q into a predefined multi-layer feed-forward neural network, whose output of last hidden layer contains n neurons. The corresponding output of the hidden layer can be denoted as $H_i^q = \left[h_{i,1}^q, \cdots, h_{i,n}^q\right]$, the output of the last layer is $O_i^q = \left[o_{i,1}^q, \cdots, o_{i,Q}^q\right]$. We denote the whole training as $O_i^q = f\left(\boldsymbol{W}_j, \mathbf{X}_i\right)$, where f is the map function and $\boldsymbol{W}_j$ is the overall parameters of the network. The connection weight from the last hidden layer to the output layer is

$\mathcal{W}_j = [\mathcal{W}_1, \mathcal{W}_2, \cdots, \mathcal{W}_Q]$. At each iteration t, the back-propagation algorithm is used to calculate the gradient of loss $\nabla L_{SML}\left(\boldsymbol{W}_j(t)\right)$ subject to $\boldsymbol{W}_j\left(t\right)$, where $L_{SML}\left(t\right)$ is a class imbalanced self-monitor loss function. Additionally, we denote the global parameters in FL server as $\boldsymbol{W}_G\left(t\right)$.

B. PPFL network architecture

In order to prevent the private diagnostic data being illegally accessed by eavesdroppers and indirectly leaked via parameter updates while training, we design a PPFL network, which contains the following six steps:

1) Access authorization: To ensure the security of private diagnostic data, biometric face recognition technology is employed. The whole biometric authentication process can be divided into two stages: enrollment, and recognition. In the first stage, the biometric template data describing the biometric characteristics of all operators is collected and constructed by the biometric collector and stored in local biometric database. In the recognition stage, the newly acquired personnel biometric data is compared with the enrolled one, and a matching score is generated. Then, the matching score is used by the system to judge whether the user currently trying to access is a legitimate one. The communication between the central server and each local client can be carried out only under the authorization of their legitimate personnel. In this case, only the authorizing central server can access local clients for FL, thus defending against malicious access by eavesdroppers.

2) Model initialization: First, we build the same network structure for the global model of the FL server and the local model of each client. Then, the global parameters $\boldsymbol{W}_G\left(1\right)$ can be obtained by random initialization. Meanwhile, in the absence of global information, we assume that the classes are in global balanced and adopt the cross entropy (CE) loss with the form

$$L_{SML}(1) = -p\log(O), \tag{1}$$

where p is the class label, O is the predicted probability.

3) Parameters broadcast: The FL sever broadcasts the global parameters and L_{SML} to each client j. Then the local model can update its own parameters $\boldsymbol{W}_j\left(t\right)$ and the loss function by

$$\left(\boldsymbol{W}_j(t), L_j(t)\right) \leftarrow \left(\boldsymbol{W}_G(t), L_{SML}(t)\right). \tag{2}$$

4) Decentralized training: Focusing on the private diagnostic data D_j in any client j, the mini-batch gradient descent algorithm is used for decentralized learning. Specifically, we first randomly divide the data D_j into B_j parts with same size Nb_j, and the loss of each part b can be obtained by

$$L_{j,b}(t) = \frac{1}{Nb_j}\sum_{i=1}^{Nb_j} L_j\left(y_i, f\left(\boldsymbol{W}_{j,b}(t), X_i\right)\right). \tag{3}$$

Then, the parameter $\boldsymbol{W}_{j,b}(t)$ is updated by

$$\boldsymbol{W}_{j,b+1}(t) \leftarrow \boldsymbol{W}_{j,b}(t) - \eta_j \nabla L_{j,b}(t), \tag{4}$$

where η_j is the learning rate of client j, Such updates are repeated B times in each round, and the training of client j will stop at a predefined round $N_{j,round}$. The final parameters $\boldsymbol{W}_{j,B}(t)$ will be regarded as well-trained parameters $\boldsymbol{W}_j^*(t)$.

5) Parameters mask and upload: Before being sent to the FL server, two parameter mask strategies are integrated into the trained parameters $\boldsymbol{W}_j^*(t)$ to prevent them from being leaked for information inference. First, we introduce the method of [24] and [32] to our proposed framework by adding Gaussian noise. The strategy for Gaussian noises mechanism can be formulated as

$$\hat{\boldsymbol{W}}_j^*(t) = \boldsymbol{W}_j^*(t) + \lambda \boldsymbol{W}_j^{Noise}(t), \tag{5}$$

where $\boldsymbol{W}_j^{\text{Noise}}(t)$ is a set of noise with mean 0 and scale of $\boldsymbol{W}_j^*(t)$ (i.e., $\sigma^2 = S^2\left(\boldsymbol{W}_j^*(t)\right)$). λ controls the ratio of noise. For each local model updated in parallel, the overall parameters of the system S can be expressed as $\hat{\boldsymbol{W}}^*(t) = \left\{\hat{\boldsymbol{W}_0}^*(t), \hat{\boldsymbol{W}_1}^*(t), \cdots, \hat{\boldsymbol{W}_C}^*(t)\right\}$. Then, a logic unit is adopted for random sampling of the overall parameters $\hat{\boldsymbol{W}}^*(t)$. The selected subset $\hat{\boldsymbol{W}}_{sub}(t)$, where $\hat{\boldsymbol{W}}_{sub}(t) \subseteq \hat{\boldsymbol{W}}^*(t)$, will be uploaded for global updates.

6) Global parameter aggregation: The FL server aggregates the selected $\hat{\boldsymbol{W}}_{sub}(t)$ by federated averaging (FedAvg), and the update of the global parameters can be denoted as

$$\boldsymbol{W}_G(t+1) = \frac{1}{M}\sum_{k=1}^{M} \boldsymbol{W}_{sub,k}(t), \tag{6}$$

where M and $\hat{\boldsymbol{W}}_{sub,k}(t)$ denote the cardinality and element of the set $\hat{\boldsymbol{W}}_{sub}(t)$, respectively.

C. Self-monitor scheme for class imbalance

To address the global class imbalance issue, a self-monitoring scheme is designed to infer the class distribution of decentralized systems with uploaded $\hat{\boldsymbol{W}}_{sub}(t)$. It is worth noting that the monitor only pays attention to the connection weight from the last hidden layer to the output layer of the global model $\mathcal{W}_G$. Assuming that instances in the same class q induce similar O^q, their gradient should be very similar, which can be denoted as

$$\nabla L_{j,b,i}^q = \overline{\nabla L^q}, \tag{7}$$

where $\nabla L_{j,b,i}^q$ is the gradient of class q with respect to the i-th instance $\boldsymbol{X}_i$ in batch b for any client j, and $\overline{\nabla L^q}$ denotes the average value of the gradient of class q. In this case, the gradient induced by class q in one global epoch is

$$\begin{aligned}\Delta_{\text{global}}\mathcal{W}_q &= \frac{1}{M}\sum_{j=1}^{M}\left[\left(-\frac{\eta_j}{Nb_j}\sum_{i=1}^{n_{j,b}^q}\nabla L_{j,b,i}^q\right)\cdot B_j\cdot N_{j,round}\right], \\ &= -\frac{\eta}{M}\overline{\nabla L^q}\left(\sum_{j=1}^{M}\frac{\eta_j}{Nb_j}n_{j,b}^q\cdot B_j\cdot N_{j,round}\right)\end{aligned} \tag{8}$$

where $n_{j,b}^q$ represents the number of instances for class q of batch b in client j. In the framework of FL, the values of η_j, Nb_j, and $N_{j,round}$ are usually the same in any client, i.e.,

$\eta = \eta_1 = \cdots = \eta_M$. Thus, we can rewrite the **Eq.** (8) as

$$\begin{aligned}\Delta_{global}\mathcal{W}_q &= -\frac{\eta}{M \cdot Nb} \cdot N_{round} \cdot \overline{\nabla L^q} \cdot \sum_{j=1}^{M} \left(n_{j,b}^q \cdot B_j \right) \\ &= -\frac{\eta}{M \cdot Nb} \cdot N_{round} \cdot \overline{\nabla L^q} \cdot \sum_{j=1}^{M} \left(N_j^q \right)\end{aligned}, \tag{9}$$

where $N_j^q = n_{j,b}^q \cdot B_j$ denotes the number of instances of class q in client j. **Eq.** (9) indicates that if we obtain the $\overline{\nabla L^q}$ or $\overline{\Delta \mathcal{W}_q}$ ($\overline{\Delta \mathcal{W}_q} = -\eta \overline{\nabla L^q}$), the global instance number ($N^q = \sum_{j=1}^{M} \left(N_j^q \right)$) can be inferred properly.

Leveraging this finding of **Eq.** (9), a monitor shown in Fig. 2 is able to estimate the class ratio of global instances through a small size auxiliary data D_a, which is voluntarily shared by any client or actively synthesized by a credit model. In data D_a, each class (q_a) has the same number of instances, denoted as $n^{aux} = n^{q_a}$. When the instances of each class q_a are fed into the global model solely, its weight update ΔW_{q_a} under n^{q_a} instances can be obtained by the monitor. Then, the average weight update generated by each q_a instance can be obtained by $\overline{\Delta \mathcal{W}_{q_a}} = \overline{\Delta \mathcal{W}_q} = \frac{\Delta \mathcal{W}_{q_a}}{n^{q_a}}$. Therefore, the global weight updates can be formulated as

$$\begin{aligned}&\Delta \mathcal{W}_G(t+1) = \sum_{q=1}^{Q} \Delta_{\text{global}} \mathcal{W}_q \\ &= \frac{N_{\text{round}}}{M \cdot Nb} \sum_{q=1}^{Q} \left(\overline{\Delta \mathcal{W}_q} \cdot N^q \right) \\ &= \frac{N_{\text{round}}}{M \cdot Nb} \sum_{q_a=1}^{Q} \left(\frac{\Delta \mathcal{W}_{q_a}}{n^{q_a}} \cdot N^q \right) \\ &= \frac{N_{\text{round}}}{M \cdot Nb} (\underbrace{\frac{\Delta \mathcal{W}_{q_a}}{n^{\text{aux}}} N^q}_{\text{generated by } q} + \underbrace{\frac{\left(\sum_{p_a=1}^{Q} \left(\Delta \mathcal{W}_{p_a} \right) - \Delta \mathcal{W}_{q_a} \right)}{n^{\text{aux}} \cdot (Q-1)} \left(\sum_{p=1}^{Q} N^p - N^q \right)}_{\text{generated by other classes}})\end{aligned} \tag{10}$$

where $\frac{N_{\text{round}}}{M \cdot Nb}$ is a constant, denoted as $Const$. $\frac{\left(\sum_{p_a=1}^{Q} (\Delta \mathcal{W}_{p_a}) - \Delta \mathcal{W}_{q_a} \right)}{n^{\text{aux}} \cdot (Q-1)} = \overline{\Delta \mathcal{W}_{others}}$ can be regarded as the average weight update generated by a non-q instance. $\sum_{p=1}^{Q} N^p - N^q$ is the total number of non-q instances.

In **Eq.** (10), expect for N^q and $\sum_{p=1}^{Q} N^p$, all the other variables are known. Therefore, **Eq.** (10) can be rewritten as

$$\begin{aligned}\Delta \mathcal{W}_G(t+1) &= Const \cdot \left(\overline{\Delta \mathcal{W}_{q_a}} \cdot N^q + \overline{\Delta \mathcal{W}_{\text{others}}} \cdot \left(\sum_{p=1}^{Q} N^p - N^q \right) \right) \\ &= Const \cdot \left(\left(\overline{\Delta \mathcal{W}_{q_a}} - \overline{\Delta \mathcal{W}_{\text{others}}} \right) N^q + \overline{\Delta \mathcal{W}_{\text{others}}} \sum_{p=1}^{Q} N^p \right).\end{aligned} \tag{11}$$

Then, the overall sample size of class q in all clients can be estimated by

$$N^q = \frac{\frac{\Delta \mathcal{W}_G(t+1)}{\text{Const}} - \overline{\Delta \mathcal{W}_{\text{others}}} \cdot \sum_{p=1}^{Q} N^p}{\left(\overline{\Delta \mathcal{W}_{q_a}} - \overline{\Delta \mathcal{W}_{\text{others}}} \right)}. \tag{12}$$

Based on **Eq.** (12), when all clients upload their sample size ($\sum_{p=1}^{Q} N^p$) in the t-th global training, the monitor can estimate the number of instances of each class $[\hat{N}^1, \cdots, \hat{N}^Q]$ in the decentralized system. Using the reverse weighting strategy of sample frequency, the class imbalance problem can be solved effectively. The sample size uploaded by each client may have a small privacy cost, but we believe it is affordable to effectively address the class imbalance in decentralized system. The evidence from [33] shows that sharing only the total sample size across all classes ($\sum_{p=1}^{Q} N_j^p$) is much less risky than sharing the sample size of each class (N_j^p), and the privacy can be protected by secure aggregation.

Then, the class imbalanced ratio of q can be defined as

$$\hat{R}^q = \frac{\hat{N}^q}{\left(\frac{\sum_{p=1}^{Q} N^p - \hat{N}^q}{Q-1} \right)}. \tag{13}$$

After careful calculation of the imbalanced ratio of each class with **Eq.** (13), the monitor can obtain the overall class imbalanced vector at the t-th global training (denoted as $\hat{R}_t = [\hat{R}^1, \cdots, \hat{R}^Q]$). We follow the self-custom loss function in [31], and design a self-monitoring loss function as

$$L_{SML}(t+1) = -\left(1 + \alpha \hat{R}_t \right) \cdot p \cdot \log(O), \tag{14}$$

where α is a hyper-parameter with default value of 0.1 in this study. After that, L_{SML} and the updated global parameters $\boldsymbol{W}_G$ will be broadcasted to each client for the discretization training of the next epoch.

D. Algorithm

We present the solution to the proposed framework in **Algorithm 1**.

Algorithm 1: The pseudo-code of the proposed CI-PPFL.

```
Input: Number of clients ($C$), training set ($[D_1, \cdots, D_C]$), maximum
       epochs ($E_{ep}$), auxiliary data ($D_a$), fraction of noise ($\lambda$), upload
       fraction of local weight ($\mu$), the number of local training rounds
       ($N_{round}$), learning rate for local update ($\eta$), local batch size
       ($Nb$).
Output: Learned parameters ($\boldsymbol{W}_G, L_{SML}$)
Authorized FL server ← Identify access authorization server with
biometric authentication
Processed $[D_1, \cdots, D_C]$ ← Pre-process $[D_1, \cdots, D_C]$ by fast
Fourier transform and reshaping
Net ← Construct a CI-PPFL framework
$(\boldsymbol{W}_G(1), L_{SML}(1))$ ← Initialize global weight $\boldsymbol{W}_G(1)$ and loss
function $L_{SML}(1)$
for $t \in [1, E_{ep}]$ do
    for each client $j \in [1, 2, \cdots, C]$ parallelly do
        $B_j$ ← split $D_j$ to same size $Nb$
        for each round $r \in [1, 2, \cdots, N_{round}]$ do
            for $b \in B_j$ do
                $\boldsymbol{W}_j(t) \leftarrow \boldsymbol{W}_j(t)$ - $\eta \nabla L_{j,b}(t)$
            end
        end
        $\hat{\boldsymbol{W}}_j^*(t)$ ← add $\lambda \boldsymbol{W}_j^{Noise}(t)$ to $\boldsymbol{W}_j^*(t)$
        $\hat{\boldsymbol{W}}_{sub}(t)$ ← random select $[\hat{\boldsymbol{W}}_0^*(t), \cdots, \hat{\boldsymbol{W}}_C^*(t)]$ at the
        ratio of $\mu$
    end
    $\boldsymbol{W}_G(t+1)$ ← aggregate $\hat{\boldsymbol{W}}_{sub}(t)$ with Eq. (6)
    for instances of class $q_a \in D_a$ do
        $\Delta \mathcal{W}_{q_a}$ ← calculate the weight change caused by $q_a$
    end
    $\left[\hat{R}^0, \cdots, \hat{R}^Q\right]$ ← measure imbalanced ratio with Eq. (13)
    $L_{SML}(t+1)$ ← update $L_{SML}(t)$ with Eq. (14)
    $(\boldsymbol{W}_G(t+1), L_{SML}(t+1))$ ← broadcast updated parameters for
    local training
end
```

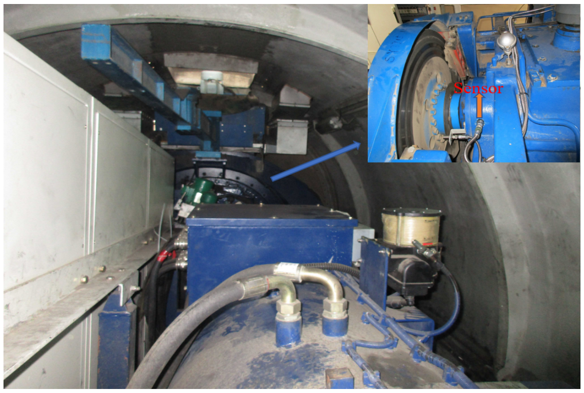

Fig. 3: Panoramic view of wind turbine nacelle and horizontally installed sensor.

IV. Experiments

A. Experimental Data and Setup

1) Data: The raw vibration data are collected from five biometric authentication wind farms located in Kangbao, Hebei Province, China. The vibration sensors shown in Fig. 3 are installed horizontally at the bearing end of a 1.5 MW doubly-fed induction generator with the sampling rate of 20 kHz. Each 0.2-second snapshot contains 4096 points. The snapshot is repeated every 2 h. During the 5-year monitoring of wind turbine bearings in five wind farms, a total of 213 run-to-failure trajectories are detected, including 122 bearing outer race faults (BOF), 45 bearing inner race faults (BIF), 23 bearing cage faults (BCF), and 23 bearing rolling element faults (BRF). In each kind of fault trajectories of any wind farm, we randomly select 1000 snapshots for this study. Finally, 1000 normal instances are also randomly selected in each client to form our overall dataset together with selected fault instances.

Before the data is fed into the CI-PPFL method for training, two pre-processing procedures are conducted. First, the fast Fourier transform technique is used to convert the signal in each snapshot from time domain to frequency domain. Owing to the symmetry of the spectrum, the first 2048 points of each spectrum form an instance. Second, the instance is reshaped into a 32x64 matrix to increase diagnosis efficiency.

2) Parameter settings: In our experiments, both the global and local models adopt the standard ResNet18 [34] as backbone. It contains 4 blocks and the number of input channels is 64. In each block, all kernel sizes are [3x3]. the stride is 1. Before the fully connected layer, an average pooling layer with the kernel size [2x2] is used to reduce the number of parameters. Then, after all the pooling results are flattened and connected to a dense layer with 512 neurons, it will be fed into the output layer with 1 neuron to diagnose the type of bearing fault. The adaptive gradient algorithm (Adagrad) is used to optimize the local model. The auxiliary data is composed of 64 instances randomly selected from each class in the overall dataset. Some of crucial hyperparameters are shown in **Table** I. We randomly collect 10% of the data from each client for the performance test of the CI-PPFL framework, and the rest for the training of each local model. Specifically, in the local training phase, we randomly split the training data of each client into training set and validation set with the proportion of 4:1. This local training process is supposed to take 20 rounds, but to improve the training efficiency and avoid over-fitting, an early stopping trick is adopted. That is, in the case where there is no improvement for 5 consecutive times, the training is terminated. Without losing generality, we successively select [BCF], [BCF, BRF], [BCF, BRF, BIF] and [BCF, BRF, BIF, BOF] as minority classes, and the remainders as majority classes. Given the randomness of network initialization, all reported results are the average of five independent trials.

In training CI-PPFL, the PyTorch framework is used. All Python codes are run on a Genuine Intel (R) (two-core 1.8GHz) processors with 32 GB RAM and NVIDIA GeForce GT 720 installed.

TABLE I: The values of crucial hyperparameters.

Hyperparameter	value	Hyperparameter	value
maximum epochs	60	Local training rounds	20
Local Batch size	32	Fraction of noise	0.01
Fraction of upload	0.8	Learning rate	10^{-3}

3) Performance metrics: We appraise the performance of the proposed CI-PPFL method with three metrics: multiclass area under curve (AUC), the accuracy of overall classes (Acc.o), the accuracy of minority classes (Acc.m). First, as the most popular performance metric in class-imbalance learning, the AUC varies the decision boundary to generalize all possible trade-offs between the false and true positive rates of a classifier, which is defined as

$$AUC = \frac{1}{2} \cdot \sin \frac{2\pi}{Q} \cdot \left(\left(\sum_{i=1}^{Q-1} r_i \cdot r_{i+1} \right) + r_Q \cdot r_1 \right), \quad (15)$$

where $r_1, r_2, \cdots, r_Q$ are pairwise AUC scores calculated by varying the classes as one versus all, and the values are sorted as $r_1 \leq r_2 \leq \cdots \leq r_Q$, Q represents the total number of classes. Acc.o and Acc.m represent the accuracy of the overall and minority classes, respectively. They can directly reflect the classification performance of the model for critical classes with fewer instances, and are in general defined as

$$Acc.o = \frac{\sum_{i=1}^{Q} \text{True Positive}_i}{\sum_{i=1}^{Q} (\text{True Positive}_i + \text{False Positive}_i)}, \quad (16)$$

$$Acc.m = \frac{\sum_{i=1}^{K} \text{True Positive}_i}{\sum_{i=1}^{K} (\text{True Positive}_i + \text{False Positive}_i)}, \quad (17)$$

where K is the number of minority classes.

To quantitatively evaluate parameter privacy, two commonly used metrics, called normalized mutual information (NMI) [35] and variance difference score (VarScore) [36], are adopted. NMI is a normalization of the Mutual Information (MI) score to scale the results between 0 and 1. Its definition is given by

$$\text{NMI}(P, Z) = 2 \cdot \frac{I(P; Z)}{H(P) + H(Z)}, \quad (18)$$

where P and Z represent the original attribute and distorted attribute of the signal, respectively. $H(\cdot)$ is the information entropy, and $I(P;Z)$ is the MI between P and Z. The closer NMI is to zero, the higher level of privacy preserving provided by Gaussian noises mechanism.

VarScore measures the variance difference between P and Z, which is defined as

$$\mathrm{VarScore} = \frac{\mathrm{Var}(P - Z)}{\mathrm{Var}(P)}. \tag{19}$$

For this metric, the larger the VarScore, the higher level of privacy preserving.

B. Comparison with benchmark methods

To illustrate the effectiveness of the proposed CI-PPFL method in addressing class-imbalance, we compare it with five benchmark methods, including the standard PPFL with cross entropy loss (denoted as PPFL-CE) [11], two data-level methods called random oversampling based PPFL (PPFL-ROS) and synthetic minority over-sampling technique based PPFL (PPFL-SMOTE) [2], and two algorithm-level methods called PPFL with focal loss (PPFL-Focal) [23] and PPFL with gradient harmonizing mechanism (PPFL-GHM) [37]. Note that the PPFL-CE method does not contain any imbalance classification technique, and serves as the baseline method. Both two data-level methods manually rebalance training sample through data manipulation, which imposes computational burden. Both PPFL-Focal and PPFL-GHM concerns minority classes by modifying the loss function, and the latter is regarded as a frontier benchmark method. To illustrate the performance of each model in dealing with imbalanced data clearly, our comparative experiments are carried out on our industrial data with four imbalanced ratios (IR) with 1:100, 1:20, 1:10, and 1:5, respectively.

The comparison results are listed in **Table** II. It can be seen that when the situation of class imbalance is getting serious, the performance of each method degrades. Moreover, the PPFL-CE method performs the worst since it lacks the corresponding technique to deal with class-imbalance. Both PPFL-ROS and PPFL-SMOTE can indeed improve the diagnostic performance by adding the number of instances of minority classes, especially when the training set is seriously out of balance. Regarding algorithm-level benchmark methods, PPFL-GHM performs better than PPFL-Focal. Among all the compared methods, the proposed CI-PPFL method is the prominent, especially in the classification of minority classes.

Statistical tests, such as the Diebold Mariano (DM) test and Wilcoxon Paired Signed-Rank (Wilcoxon) test, can provide evidence to the claim that the CI-PPFL is superior to the five benchmark methods. Specifically, we pair CI-PPFL (M1) with each of the five benchmark models (M2) and conduct the two tests. The argument 'alternative' in both two tests is set to be 'less'. In other words, the alternative hypothesis is that M2 is less accurate than M1. In **Table** II, DM test and Wilcoxon test are conducted at the 10% significance level, and the entries which are significantly worse than the proposed CI-PPFL are marked as * and ‡, respectively. In almost all the cases, the proposed CI-PPFL method significantly outperforms PPFL-CE, PPFL-ROS, and PPFL-Focal. Meanwhile, CI-PPFL is significantly better than PPFL-SMOTE when IR is 1:5.

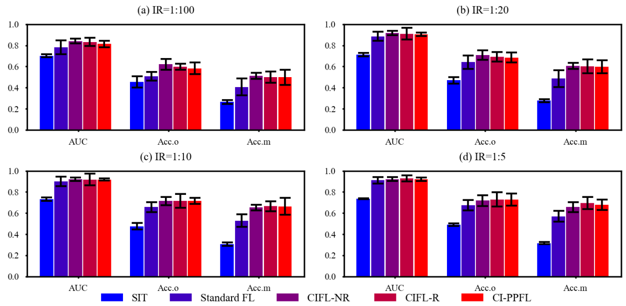

Fig. 4: The comparison of performance on different variants.

It also has a significant advantage in addressing minority classes compared to PPFL-GHM when the sample is seriously imbalanced.

C. Ablation Study

To evaluate the role of each component of the proposed CI-PPFL, an ablation study is conducted. Specifically, we consider the following model variants.

- Self-isolation training (SIT). This variant only trains the network based on the local data within one client, and clients are completely isolated from each other.
- Standard FL. The standard FL method that ignores class-imbalance and privacy leakage caused by gradient traceability.
- CIFL-NR. Our class-imbalanced self-monitor scheme is integrated into the standard FL, without the mechanisms of Gaussian noise and parameter ratio update.
- CIFL-R. This variant adds a Gaussian noise mechanism to the CIFL-NR.

For a fair comparison, all variants use ResNet18 as the backbone, with the same parameter settings as the aforementioned CI-PPFL. We show the evaluation results in Fig. 4, where the error bar represents the standard deviation. Some interesting findings emerge. First, compared with the self-isolation training method, the FL based methods can improve the diagnostic performance through parameter sharing and aggregation. Second, the self-monitor scheme for class-imbalance can significantly improve the accuracy of FL for minority classes without poisoning the overall performance. Third, we note that when IR is 1:100, the CIFL-NR is slightly superior to the CIFL-R. It may indicate that adding Gaussian noise to the training parameters for the purpose of privacy protection results in the diagnostic performance a slight decline, when the sample is seriously out of balance. This phenomenon is reversed when IR is 1:10 and 1:5. In addition, comparing the results of CI-PPFL and CIFL-R, we find that their difference is not obvious except when IR=1:100, the accuracy of CI-PPFL is slightly lower than that of CIFL-R.

D. Sensitivity Analysis

In this section, we conduct a sensitivity analysis for several crucial parameters, including the pre-processing methods, the

TABLE II: Performance comparison among various methods on the real-world industrial data.

Model	IR=1:100			IR=1:20			IR=1:10			IR=1:5		
	AUC (%)	Acc.o (%)	Acc.o (%)	AUC (%)	Acc.o (%)	Acc.o (%)	AUC (%)	Acc.o (%)	Acc.o (%)	AUC (%)	Acc.o (%)	Acc.o (%)
PPFL-CE	73.99*‡	52.14*‡	43.27*‡	80.17*‡	56.90*‡	53.49*‡	82.84*‡	58.87*‡	54.49*‡	83.25*‡	57.89*‡	54.63*‡
PPFL-ROS	**88.82**	62.00	46.00*‡	88.94	67.49	58.35*‡	91.41	70.89‡	64.83*‡	91.55*‡	71.91*‡	66.18*‡
PPFL-SMOTE	85.80	**64.52**	49.73	88.78‡	67.17	**59.89**	91.56	**71.56**	65.72	91.85‡	72.06*‡	67.34*‡
PPFL-Focal	74.85*‡	50.21*‡	47.61*‡	85.64*‡	61.18*‡	55.66*‡	83.54*‡	62.49*‡	59.24*‡	83.90*‡	61.02*‡	58.18*‡
PPFL-GHM	84.45	58.72	49.08*‡	88.75‡	67.85	59.30*‡	90.63	71.39	65.43‡	91.63*‡	72.26	67.72
CI-PPFL	81.48	58.25	**49.89**	**90.51**	**68.56**	**59.89**	**91.77**	71.50	**66.45**	**92.22**	**72.64**	**67.85**

Notes: boldface represents better result in each column. * and ‡ indicate that the difference between marked method and the proposed method is statistically using DM test and Wilcoxon test at the 10% significance level, respectively.

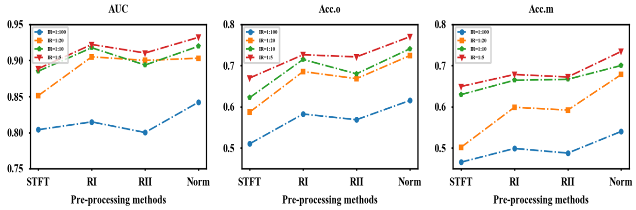

Fig. 5: Sensitivity analysis of the pre-processing methods.

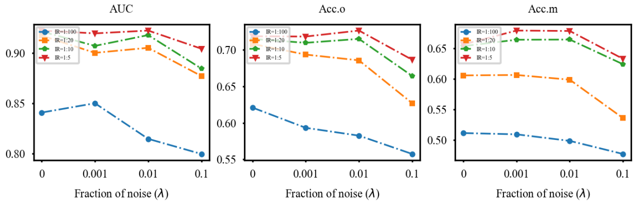

Fig. 7: Sensitivity analysis of the fraction of noise (λ).

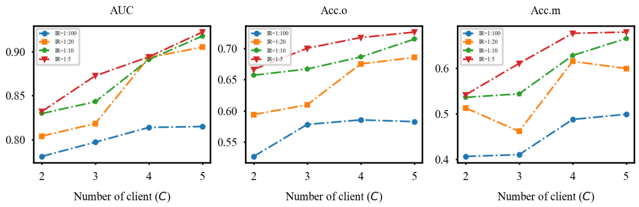

Fig. 6: Sensitivity analysis of the number of clients (C).

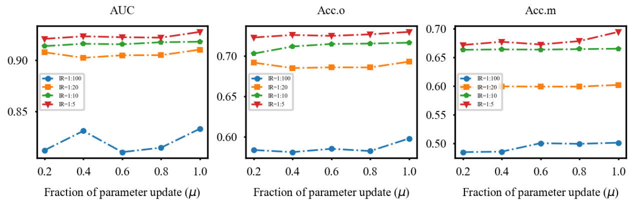

Fig. 8: Sensitivity analysis of the upload fraction of local weight (μ).

number of client (C), the fraction of noise (λ), the fraction of parameter update (μ), the learning rates, the batch sizes, and the backbone networks.

1) Pre-processing methods: Fig. 5 shows the comparison results across four data pre-processing methods: (a) convert the raw signal into two-dimensional time-frequency domain signal through Short-time Fourier transform (STFT); (b) reshape the frequency domain signal processed by fast Fourier transform into a 32x64 matrix (denoted as RI), and a similar trick can be found [38]; (c) reshape the frequency domain signal into a 64x32 matrix (denoted as RII); and (d) normalize the two-dimensional matrix generated by RI (denoted as Norm). It can be seen that STFT is inferior to the other three methods. One possible reason is that it is difficult to fully extract vibration information of variable frequency with a fixed-length window. Methods RI and RII have similar diagnostic performance, indicating no significant difference between the various reshaping methods. Normalization brings a certain performance improvement in this study. However, it requires each client to share the maximum and minimum values of its vibration data, which has the risk of privacy leakage.

2) Number of clients (C): Fig. 6 shows the performance of the proposed CI-PPFL method under different client numbers. It is obvious that the diagnostic accuracy of CI-PPFL does improve with the increase of client numbers. It indicates that the proposed framework has the ability to learn the decentralized knowledge of wind turbine, even though the local data cannot be communicated and shared.

3) Fraction of noise (λ): Fig. 7 illustrates the effect of noise parameter λ, which controls the ratio of ground-truth signal to random noise. Small λ can maintain the excellent diagnostic performance of the proposed framework. However, when $\lambda = 0.1$, the three metrics dramatically drop. It shows that the added noise is too loud and harms the proposed framework. Empirically, we recommend $\lambda = 0.01$ for the CI-PPFL method.

4) Upload fraction of local weight (μ): Fig. 8 shows that the upload fraction of local weight μ has little effect on the performance of the CI-PPFL. To make a trade-off between privacy protection and iterations, we choose $\mu = 0.8$ in the proposed framework.

5) Batch sizes: Fig. 9 shows that as the batch size increases, the performance of the proposed framework degrades on all four types of imbalanced data. This means that when the batch size is small, it may be more suitable for fine-tuning the model parameters, but it brings the burden of increasing the number of iterations. In this study, when the batch size is 32, a good trade-off between computational efficiency and accuracy of the proposed framework can be achieved.

6) Learning rates: Fig. 10 shows that the difference in the overall performance of the proposed CI-PPFL with four different learning rates is very small. Only when the learning rate is 0.0001, the proposed framework gets slightly worse. One possible reason is that the learning rate is too small to get rid of the suboptimal solution. Empirically, we recommend

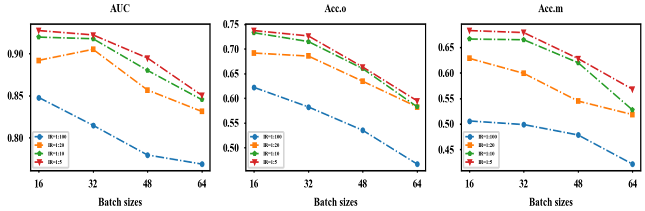

Fig. 9: Sensitivity analysis of the batch sizes.

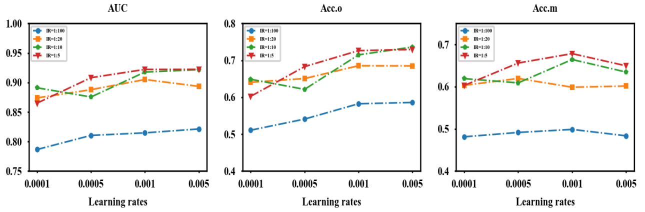

Fig. 10: Sensitivity analysis of the learning rates.

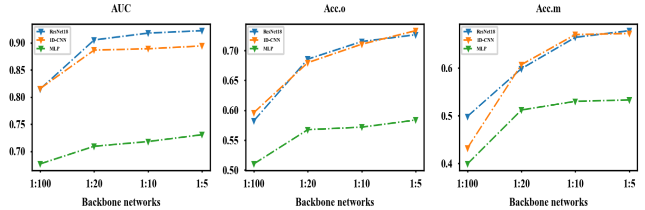

Fig. 11: Sensitivity analysis of the backbone networks.

the learning rate of the proposed framework is 0.001.

7) Backbone networks: The performance comparison between the shallow feed-forward network, namely multilayer perceptron (MLP), one-dimensional convolutional neural network (1D-CNN) [39], and the deep ResNet18 as the backbone network is shown in Fig. 11. As we can see, the ResNet18-based framework significantly outperforms the MLP-based ones. Except for the Acc.m metric with imbalanced ratios of 1:20 and 1:10, the performance of ResNet18 is similar to or better than 1D-CNN as well. We recommend the feed-forward network with deep feature extraction as the backbone network in CI-PPFL, e.g., ResNet18.

E. Security Analysis

In this section, we first quantitatively evaluate the security of parameter privacy by different fraction of Gaussian noise (λ). The comparison results are shown in Fig. 12. It can be seen that under different imbalanced ratios, the NMI decreases continuously with the increase of λ. This indicates that the more Gaussian noise added, the higher the level of parameter privacy protection. The results from VarScore also support this finding. In addition, we note that when $\lambda = 0.1$, the Gaussian noise mechanism makes the uploaded parameters have a large variance with the original parameters, but may sacrifice a certain prediction accuracy (see Fig. 7).

Additionally, to analyze the security more thoroughly, we also conduct a qualitative comparison between the proposed framework and the existing central training [9], conventional FL [15], and DP-FL [11] in three security attributes, including data privacy, parameter privacy and anti-malicious access. The comparison results are presented in Table III. It shows that the FL-based framework is able to preserve data privacy compared with central training. Differential privacy technique can avoid the parameter privacy risk, but there is still the risk of malicious access. In contrast, our proposed framework uses biometric authentication technique, which only authorizes legitimate entities to access private data and defend against malicious attacks, thus providing more comprehensive security.

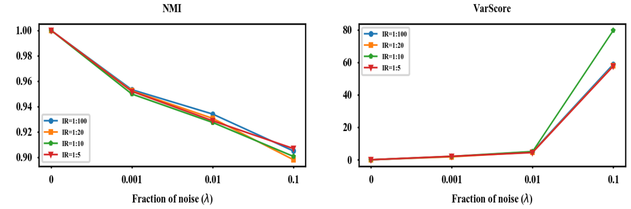

Fig. 12: Security analysis of the fraction of noise (λ).

TABLE III: Security comparison among various learning frameworks.

Frameworks	Data privacy	Parameters privacy	Anti-malicious access
Central training	No	No	No
FL	Yes	No	No
DP-FL	Yes	Yes	No
Proposed framework	Yes	Yes	Yes

V. Conclusion

A class-imbalance privacy-preserving federated learning framework, namely CI-PPFL, for fault diagnosis of decentralized wind turbine is proposed. FL firstly learns the decentralized knowledge of isolated biometric authentication clients without data sharing, thus alleviating the risk of privacy leakage. Then, two privacy-enhancing techniques, the gradient noise mechanism and parameter proportional update strategy, are designed to hinder gradient tracking and information inference. Besides, a gradient self-monitor scheme is integrated into the FL to acknowledge the global imbalance information for class-imbalance fault diagnosis.

The proposed CI-PPFL method is evaluated on a real-world industrial wind turbine dataset. Its superiority in addressing class-imbalance issue is validated by compared with several state-of-the-art methods. Meanwhile, the Wilcoxon test and DM test make its advantage ascertain. Then, an ablation study indicates the two privacy-enhancing techniques will not harm the diagnostic performance. After that, the sensitivity analysis details the impact of crucial parameters.

Future research in partial federated transfer learning for imbalance issue is warranted. The proposed CI-PPFL just focus on the case that the label spaces of each client are identical. The proposed framework may not work properly when encountering heterogeneous label subspaces. The partial domain adaptation technique has the ability to reduce the space discrepancy of different clients. In addition, the fault diagnosis scheme integrating wind turbine vibration data and SCADA data will become a core of our follow-up study. In this regard, these improvements can broaden the application scope of CI-PPFL in practical engineering.

References

[1] W. Y. B. Lim, N. C. Luong, D. T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, D. Niyato, and C. Miao, "Federated learning in mobile edge networks: A comprehensive survey," *IEEE Commun. Surveys Tuts*, vol. 22, no. 3, pp. 2031–2063, 2020.

[2] M. Duan, D. Liu, X. Chen, R. Liu, Y. Tan, and L. Liang, "Self-balancing federated learning with global imbalanced data in mobile systems," *IEEE Trans. Parallel Distrib. Syst.*, vol. 32, no. 1, pp. 59–71, 2021.

[3] W. Zhang and X. Li, "Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions," *Struct. Health Monit.*, p. 147592172110292, 2021.

[4] Q. Jiang, N. Zhang, J. Ni, J. Ma, X. Ma, and K.-K. R. Choo, "Unified biometric privacy preserving three-factor authentication and key agreement for cloud-assisted autonomous vehicles," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 9390–9401, 2020.

[5] T. Phillips, X. Zou, F. Li, and N. Li, "Enhancing biometric-capsule-based authentication and facial recognition via deep learning," in *Proc. 24th ACM Symposium Access Control Models Technol.*, 2019, pp. 141–146.

[6] A. K. Das, M. Wazid, N. Kumar, A. V. Vasilakos, and J. J. P. C. Rodrigues, "Biometrics-based privacy-preserving user authentication scheme for cloud-based industrial internet of things deployment," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 4900–4913, 2018.

[7] K. Zhou and J. Ren, "Passbio: Privacy-preserving user-centric biometric authentication," *IEEE Trans. Informat. Forens. Secur.*, vol. 13, no. 12, pp. 3050–3063, 2018.

[8] G. Kaissis, A. Ziller, J. Passerat-Palmbach, T. Ryffel, D. Usynin, A. Trask, I. Lima, J. Mancuso, F. Jungmann, M.-M. Steinborn, A. Saleh, M. Makowski, D. Rueckert, and R. Braren, "End-to-end privacy preserving deep learning on multi-institutional medical imaging," *Nature Machine Intelligence*, vol. 3, no. 6, pp. 473–484, 2021.

[9] Q. Xu, S. Lu, W. Jia, and C. Jiang, "Imbalanced fault diagnosis of rotating machinery via multi-domain feature extraction and cost-sensitive learning," *J. Intell. Manuf.*, vol. 31, no. 6, pp. 1467–1481, 2020.

[10] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniquespart II: Fault diagnosis with knowledge-based and hybrid/active approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, 2015.

[11] B. Zhao, K. Fan, K. Yang, Z. Wang, H. Li, and Y. Yang, "Anonymous and privacy-preserving federated learning with industrial big data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 9, pp. 6314–6323, 2021.

[12] R. Rahimilarki, Z. Gao, A. Zhang, and R. Binns, "Robust neural network fault estimation approach for nonlinear dynamic systems with applications to wind turbine systems," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6302–6312, 2019.

[13] Z. Gao and X. Liu, "An overview on fault diagnosis, prognosis and resilient control for wind turbine systems," *Processes*, vol. 9, no. 2, p. 300, 2021.

[14] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist.*, vol. 54, 2017, pp. 1273–1282.

[15] Z. Zhang, X. Xu, W. Gong, Y. Chen, and H. Gao, "Efficient federated convolutional neural network with information fusion for rolling bearing fault diagnosis," *Control Engineering Practice*, vol. 116, p. 104913, 2021.

[16] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Federated learning for machinery fault diagnosis with dynamic validation and self-supervision," *Knowledge-Based Syst.*, vol. 213, p. 106679, 2021.

[17] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, p. 13221333.

[18] Z. Zhang, C. Guan, H. Chen, X. Yang, W. Gong, and A. Yang, "Adaptive privacy preserving federated learning for fault diagnosis in internet of ships," *IEEE Internet Things J.*, pp. 1–1, 2021.

[19] L. Wang, S. Xu, X. Wang, and Q. Zhu, "Addressing class imbalance in federated learning," *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 11, pp. 10 165–10 173, 2021.

[20] Q. He, Y. Pang, G. Jiang, and P. Xie, "A spatio-temporal multiscale neural network approach for wind turbine fault diagnosis with imbalanced SCADA data," *IEEE Trans. Ind. Informat.*, vol. 17, no. 10, pp. 6875–6884, 2021.

[21] X. Cheng, F. Shi, X. Liu, M. Zhao, and S. Chen, "A novel deep class-imbalanced semisupervised model for wind turbine blade icing detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2558–2570, 2022.

[22] M. Zareapoor, P. Shamsolmoali, and J. Yang, "Oversampling adversarial network for class-imbalanced fault diagnosis," *Mech. Syst. Signal Process.*, vol. 149, p. 107175, 2021.

[23] Y. Wang, G. Gui, H. Gacanin, B. Adebisi, H. Sari, and F. Adachi, "Federated learning for automatic modulation classification under class imbalance and varying noise condition," *IEEE Trans. Cogn. Commun. Netw.*, pp. 1–1, 2021.

[24] A. Neelakantan, L. Vilnis, Q. Le, I. Sutskever, L. Kaiser, K. Kurach, and J. Martens, "Adding gradient noise improves learning for very deep networks," p. arXiv:1511.06807, 2015.

[25] A. Hard, K. Rao, R. Mathews, S. Ramaswamy, F. Beaufays, S. Augenstein, H. Eichner, C. Kiddon, and D. Ramage, "Federated learning for mobile keyboard prediction," p. arXiv:1811.03604, 2018.

[26] A. Vaid, S. K. Jaladanki, and J. Xu, "Federated learning of electronic health records to improve mortality prediction in hospitalized patients with covid-19: Machine learning approach," *JMIR Med Inform*, vol. 9, no. 1, 2021.

[27] A. Ziller, D. Usynin, R. Braren, M. Makowski, D. Rueckert, and G. Kaissis, "Medical imaging deep learning with differential privacy," *Sci Rep*, vol. 11, no. 1, p. 13524, 2021.

[28] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, p. 308318.

[29] Z. Zhou and X. Liu, "Training cost-sensitive neural networks with methods addressing the class imbalance problem," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 1, pp. 63–77, 2006.

[30] P. Domingos, "Metacost: A general method for making classifiers cost-sensitive," in *Proc. 5th ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 1999, p. 155164.

[31] S. H. Khan, M. Hayat, M. Bennamoun, F. A. Sohel, and R. Togneri, "Cost-sensitive learning of deep feature representations from imbalanced data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3573–3587, 2018.

[32] W. Zhang, H. Quan, and D. Srinivasan, "An improved quantile regression neural network for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4425–4434, 2019.

[33] A. Salem, A. Bhattacharya, M. Backes, M. Fritz, and Y. Zhang, "Updates-leak: data set inference and reconstruction attacks in online learning," in *Proc. 29th USENIX Conf. Secur. Symp.*, 2020, pp. 6302–6312.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit*, 2016, pp. 770–778.

[35] Q. Li, R. Heusdens, and M. G. Christensen, "Privacy-preserving distributed optimization via subspace perturbation: A general framework," *IEEE Trans. Signal Process.*, vol. 68, pp. 5983–5996, 2020.

[36] E. Bertino, D. Lin, and W. Jiang, "A survey of quantification of privacy preserving data mining algorithms," in *Privacy-Preserving Data Mining (Advances in Database Systems)*, vol. 5, 2008, pp. 183–205.

[37] B. Li, Y. Liu, and X. Wang, "Gradient harmonized single-stage detector," *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 01, pp. 8577–8584, 2019.

[38] T. Xie, X. Huang, and S. K. Choi, "Intelligent mechanical fault diagnosis using multisensor fusion and convolution neural network," *IEEE Trans. Ind. Informat.*, vol. 18, no. 5, pp. 3213–3223, 2022.

[39] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, 2016.

Shixiang Lu (Student Member, IEEE) received the B.S. degrees from Hefei University of Technology, China, in 2017, where he is currently pursuing the Ph.D degree with the School of Management. Since August, 2021, he has served as a researcher at the Faculty of Engineering and Environment, University of Northumbria, Newcastle upon Tyne, U.K.

His research interests include smart manufacturing, computational intelligence applications in fault diagnosis and prognosis.

Zhiwei Gao (Senior Member, IEEE) received the Ph.D. degrees in systems engineering from Tianjin University, Tianjin, China, in 1996.

He is currently the Head of Electrical Power and Control Systems Research Group, Faculty of Engineering and Environment, University of Northumbria, Newcastle upon Tyne, U.K. His research interests include data-driven modeling, estimation and filtering, fault diagnosis, fault tolerant control, resilient control, artificial intelligence and machine learning, intelligent optimization, distribution estimation and control, wind turbine systems, power electronics, electrical vehicles, bioinformatics, and healthcare systems.

Dr. Gao is an Associate Editor for the IEEE Transactions on Systems Man Cybermetics-Systems, IEEE Transactions on Industrial Informatics, IEEE Transactions on Industrial Electronics, and IEEE Transactions on Automatic Control. In addition, he was the Associate Editor of IEEE Transactions on Control Systems Technology (from 2009-2015).

Qifa Xu received the B.E. degree in the Department of Mathematics from Fuyang Normal University, China, the M.S. degree from Dongbei University of Finance and Economics, China, and the Ph. D. degree in School of Management from Tianjin University, China, in 2006. He is currently a Professor with the School of Management, Hefei University of Technology, China. He has held a visiting position of statistics at the Florida State University, U.S.A, from May, 2012 to May, 2013. His research interests include financial big data analysis, statistical learning, and smart manufacturing.

Cuixia Jiang received the B.E. degree in the Department of Mathematics from Fuyang Normal University, China, and the Ph. D. degree in School of Management from Tianjin University, China, in 2008. She is currently an Associated Professor with the School of Management, Hefei University of Technology, China. Her current research interests include financial big data analysis, financial time series analysis, and statistical learning.

Aihua Zhang received her B.Sc. degree from Jinzhou Teachers College in 2000, M.sc. degree from Bohai University in 2008, and Ph.D. degree from Harbin Institute of Technology in 2014. She is a Full Professor at College of Physical Science and Technology in Bohai University. Her current research interests include fault diagnosis, fault tolerance, and attitude control of satellites.

Xiangxiang Wang received the B.E. degree from Hefei University of Technology, and the M.Eng. degree in Department of Precision Machinery and Precision Instrumentation from University of Science and Technology of China, in 2014. She is currently pursuing the Ph.D degree with the School of Management. Her research interests include data-driven fault diagnosis and prognosis, reliability engineering, and edge computing.