Computers, Materials & Continua

Tech Science Press

Doi:10.32604/cmc.2025.075156

ARTICLE

Robust and Efficient Federated Learning for Machinery Fault Diagnosis in Internet of Things

Zhen Wu[1,2], **Hao Liu**[3], **Linlin Zhang**[4], **Zehui Zhang**[5,*], **Jie Wu**[1], **Haibin He**[1] **and Bin Zhou**[6]

[1]Ningbo C.S.I. Power & Machinery Group. Co., Ltd., Ningbo, 315020, China
[2]Zhejiang Institute of Communications, Hangzhou, 311112, China
[3]School of Software, Nankai University, Tianjin, 300350, China
[4]China Automotive Technology and Research Center Co., Ltd., Tianjin, 300300, China
[5]China-Austria Belt and Road Joint Laboratory on Artificial Intelligence and Advanced Manufacturing, Hangzhou Dianzi University, Hangzhou, 310018, China
[6]Jiangxi Tsinghua Tellhow Sanbo Electric Machinery Co., Ltd., Nanchang, 330096, China

*Corresponding Author: Zehui Zhang. Email: zhangtianxia918@163.com

Received: 26 October 2025; Accepted: 21 November 2025; Published: 10 February 2026

ABSTRACT: Recently, Internet of Things (IoT) has been increasingly integrated into the automotive sector, enabling the development of diverse applications such as the Internet of Vehicles (IoV) and intelligent connected vehicles. Leveraging IoV technologies, operational data from core vehicle components can be collected and analyzed to construct fault diagnosis models, thereby enhancing vehicle safety. However, automakers often struggle to acquire sufficient fault data to support effective model training. To address this challenge, a robust and efficient federated learning method (REFL) is constructed for machinery fault diagnosis in collaborative IoV, which can organize multiple companies to collaboratively develop a comprehensive fault diagnosis model while keeping their data locally. In the REFL, the gradient-based adversary algorithm is first introduced to the fault diagnosis field to enhance the deep learning model robustness. Moreover, the adaptive gradient processing process is designed to improve the model training speed and ensure the model accuracy under unbalance data scenarios. The proposed REFL is evaluated on non-independent and identically distributed (non-IID) real-world machinery fault dataset. Experiment results demonstrate that the REFL can achieve better performance than traditional learning methods and are promising for real industrial fault diagnosis.

KEYWORDS: Federated learning; adversary algorithm; Internet of Vehicles (IoV); fault diagnosis

1 Introduction

Fault diagnosis plays a crucial role in modern vehicles, which can improve vehicle safety and reduce maintenance costs [1–3]. Various complex machinery and equipment are used in modern vehicles, whose faults are more difficult to diagnose. Studies on fault diagnosis in the literature can generally be classified into several categories, as discussed in [4,5]. The model-based method calculates the difference between the predicted value of the mathematical physics model and the monitor value to diagnose faults. While these methods can depict the process dynamics mathematically, the mathematical physics model for complex machinery is hard to establish. By contrast, data-driven fault diagnosis methods directly use historical data. The Internet of Vehicles (IoV) [6–8] facilitates the collection and analysis of unprecedented volumes of data, thereby paving the way for the advancement of intelligent methods. Recently, DL-based diagnostic methods have been much more popular in many devices, including electrical motors, diesel engines, gearboxes, etc. [9–12].

Although DL-based fault diagnosis methods achieve excellent performance in many case studies, the model accuracy may drop when the model is applied to real environments. That is because the training data is difficult to cover the whole dynamic situation. Hence, researchers usually require collecting a large amount of data to train a high-performance model. Due to the high capital cost, it is difficult for a single vehicle manufacturer to obtain sufficient data to develop a high-performance DL-based fault diagnosis model, which hinders the DL application in the field.

In general, multiple manufacturers often equip their products with identical or similar core devices. For instance, numerous vehicle models across different brands adopt the same type of engine, enabling unified monitoring and data aggregation for these engines. Hence, to expand the dataset at lower costs, a straightforward method called centralized learning (CL) is to aggregate the local monitored data of different companies. In this method, a centralized server uses the aggregated data to develop the diagnostic model and enhance the model's performance. However, vehicle manufacturers are reluctant to share their proprietary data with others due to potential business competition concerns.

Federated Learning (FL) [13] serves as a promising method to the aforementioned issues by harnessing the local data resources of distributed participants to build a powerful DL-based model collaboratively. As presented in Fig. 1, multiple participants perform model training using their proprietary local data and transmit their updated models. The server aggregates the upload data of the different participants to update the global model. Only model parameters can be transmitted, and data resources are safely stored locally. To improve the FL efficiency, some studies [14,15] used momentum terms to accelerate model convergence speed to reduce resource costs. Most accelerating federated learning studies assumed that the data of the participants are independent and identically distributed (IID). However, in the practical fault diagnosis scenarios, the machines at different factories/companies often work under different conditions. These monitored data are subject to different distributions, which may result in model performance degradation when the FL system uses the second-order momentum term to accelerate model training speed. Moreover, the operation mechanisms of DL models are not yet available. Some studies demonstrate that the input data with noise (such as Gaussian noise) may cause the fault diagnosis model to perform poorly [16,17].

In this study, a robust and efficient federated learning approach is proposed for machinery fault diagnosis in collaborative AIoT. The primary contributions made by this study are summarized as follows:

1) FL for machinery fault diagnosis is studied in this article, which is seldom studied in the current literature. Different vehicle manufacturers can collaboratively develop a global fault diagnosis model without transferring their data to external parties.

2) To the best of our knowledge, the gradient-based adversarial algorithm is first introduced to the fault diagnosis field. The adversarial algorithm generates a lot of adversarial samples at low computation costs for enhancing the fault diagnosis model robustness to resist external noises.

3) To accelerate model convergence speed and resist the unbalanced data scenarios in federated learning tasks, we proposed a local model updating scheme based on Adam optimization method.

The rest of this article is structured as follows. Section 2 presents the related works. The preliminaries are introduced in Section 3. Section 4 presents the REFL method for machine machinery fault diagnosis in detail and experimentally evaluates the method in Section 5. Section 6 discusses the proposed method by comparing it to similar studies. Finally, Section 7 concludes this study. The notation definitions of this paper are listed in Table 1.

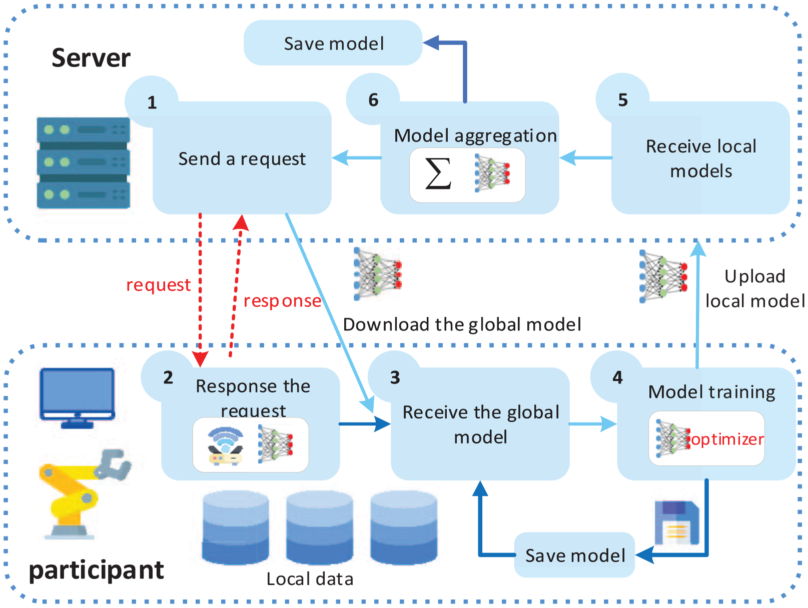


Figure 1: Illustration of the FL system

Table 1: Notation definitions

Notation	Definition
N	The industrial participant number
b_k	The mini-batch size of the participant k
D_k	The local data of the participant k
$L()$	The empirical loss function
$\mathbf{g}_{\text{local},\,k}$	The local model gradients of the participant k
$\mathbf{g}_{\text{global}}$	The global model gradients
$\mathbf{w}_{\text{local},\,k}$	The local model weights of the participant k
$\mathbf{x}$; $\mathbf{y}$	The sample; the label
$\mathbf{x}'$; ε	The adversary samples; multiplier to ensure the perturbations are small
ρ	The proportion of data used to make adversary samples
η; β_1; β_2	Learning rate; first-order momentum rate; second-order momentum rate

2 Related Work

Deep neural network models can directly extract deep-level features from raw monitoring signals, facilitating the design of end-to-end fault diagnosis methods that take raw monitoring signals as inputs and output fault diagnosis results. The emergence of deep learning significantly reduces the difficulty of building fault diagnosis models for practical tasks. Many studies [18–20] directly adopted the vibration and acoustic signals to diagnose mechanical device health conditions. Khorram et al. [18] designed an end-to-end diagnostic algorithm using deep learning neural networks, which directly uses accelerometer signals as model inputs. Ben Abid et al. [20] proposed the intelligent induction motor fault diagnosis method called deep-SincNet that can automatically extract fault features from the raw motor current.

As mentioned in the above section, current DL-based fault diagnosis methods significantly suffer from insufficient data. FL, as a promoting collaborative learning method, has been adopted in the fault diagnosis field. Zhang et al. [21] designed an efficient federated framework for rolling bearing fault diagnosis and utilized the first-order momentum term to improve model training speed. Li et al. [22] proposed a stacking model for diagnosing permanent magnet synchronous motors and used federated learning to train the model for overcoming data islanding. Zhang et al. [23] designed a federated transfer learning approach for machinery diagnostics. While these FL-based studies can alleviate the data island issue, they did not address the need to enhance the model robustness.

To improve the diagnostic performance of models in real-world industrial scenarios, the robustness of deep learning (DL) models has garnered significant attention. Numerous studies have shown that DL models are susceptible to attacks from noise-contaminated samples. Most existing studies employ generative models such as generative adversarial networks or adversarial training to enhance robustness. Ren et al. [24] proposed a Few-shot GAN that avoids the overfitting problem encountered when training GANs with very few samples. Wang et al. [25] presented a robust fault diagnosis framework based on an improved domain-adversarial neural network integrated with multi-module fusion, which enhances the robustness of the model. Wang et al. [26] proposed a traceable multi-domain collaborative generative adversarial network, aiming to improve the performance of fault diagnosis models on imbalanced data. Wang et al. [27] proposed a novel method to enhance the robustness of fault diagnosis models for high-speed trains. However, such methods based on generative adversarial networks require substantial computational resources when generating adversarial samples. Some gradient-based adversary algorithms are proposed, which can generate a lot of adversary samples for model training at low computation costs. The gradient-based adversarial methods have demonstrated significant effectiveness in enhancing model robustness across domains such as images and text. Consequently, this study pioneers the application of gradient-based adversarial algorithms in the field of fault diagnosis to enhance the fault diagnosis model robustness.

3 Preliminaries

3.1 FL

FL is adopted to solve the machine fault diagnosis task in this article. We consider a standard model of FL in which several participants cooperatively train a deep learning model, which consists of a cloud server $\mathcal{S}$ and multiple distributed participants. The participants connect with the server via secure communication channels to protect the integrity and security of the uploaded ciphertext.

1) Cloud server $\mathcal{S}$: The core task of the server is to aggregate the model parameters uploaded by participants to update the global model parameters, and then broadcast the updated parameters to all participants.

2) Participants $\mathcal{P}$: Each participant $\mathcal{P}_K$ $(k = 1, \ldots, N)$ stores a replica of the global model parameters and holds its local private data D_k. The participant trains the local model on its private data and then encrypts the local model parameters before uploading it to the server. After receiving the global model parameters from the server, a new round of training is carried out. This process is repeated until a satisfactory convergence criterion is obtained. FL's objective is to find local optimal weights $\mathbf{w}$ that simultaneously minimize the expected experience loss on all participants' data:

$$\min \sum_{i=1}^{N} L_i \left(f \left(\mathbf{x}; \mathbf{w}_{\text{global}} \right), \mathbf{y} \right) \tag{1}$$

where L represents the loss function of $\mathcal{P}_K$, respectively.

The assumptions for federated learning are as follows:

1) Multiple industrial participants hold the data of the same/similar machines for collaboratively training a fault diagnosis model.

2) The participants have different data distributions due to varying operating conditions.

3) The cloud server and participants are assumed as honest which is commonly used in the FL field. That means that they will rigorously follow the designed protocols. The data resources of each participant cannot be accessed by others.

3.2 Gradient-Based Adversary Algorithm

Szegedy [28] first defined adversarial samples as malicious inputs produced from legitimate samples by adding small perturbations to deep neural classification models into misclassifying. Literature indicated that gradient-based adversary algorithms (e.g., [29,30]) can quickly produce adversarial samples, which fool the deployed deep neural model successfully. As a famous gradient-based adversary algorithm, FGSM proposed by Goodfellow et al. [30] can fast generate adversarial examples utilizing model gradient information. Inspired by the above literature, this study uses the algorithm to generate adversary samples to enhance the model robustness. As shown in Fig. 2, an attacker adds small perturbations to the original images, which results in model misclassification. The process of adding these perturbations involves six steps in this order:

1. Taking a training sample
2. Making predictions on the sample with a trained deep-learning model
3. Computing the loss value of the prediction result based on the true class label
4. Computing the model gradients for the input sample
5. Calculating the sign of the model gradients
6. Using the signed gradients to generate the adversarial samples as $\mathbf{x}' = \mathbf{x} + \varepsilon \cdot sign\left(\nabla_x J\left(\mathbf{w}, \mathbf{x}, \mathbf{y}\right)\right)$.

3.3 Paillier Homomorphic Encryption

This paper selects the Paillier homomorphic encryption algorithm, which is briefly described as:

1) Key Generation: Key Generation generates the public key $pk(n,g)$, secret key $sk(\lambda)$ and sends them to all clients.

2) Encryption: Each edge encrypts the plaintext m using the public key $pk(n,g)$ by

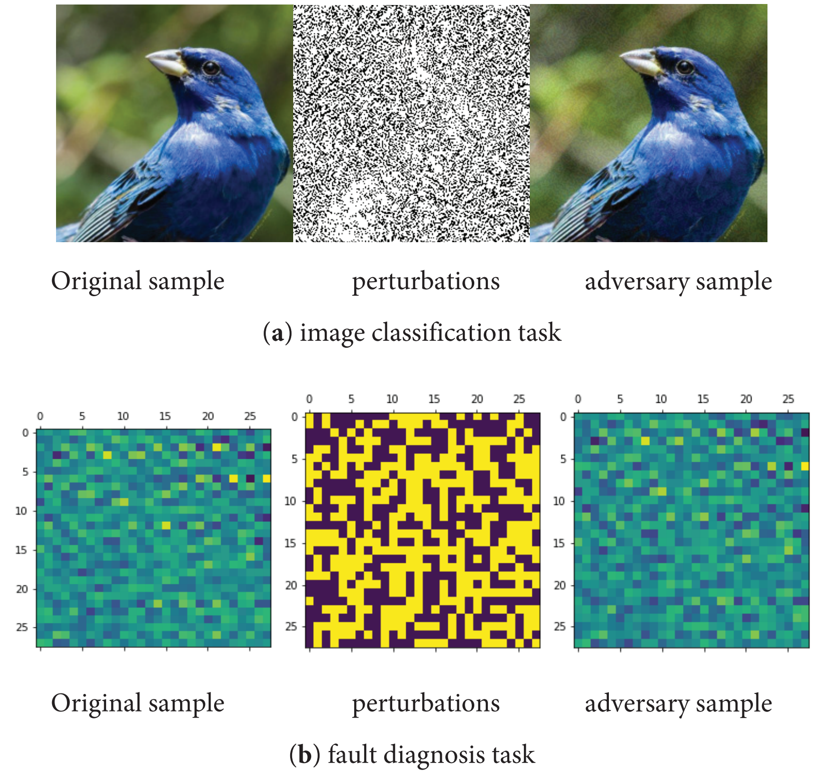
$$\mathrm{E}\left(\mathbf{w}_{\mathrm{local},k}\right) = g^{w} \cdot r^{n} \bmod n^{2} \tag{2}$$

3) Aggregation: The server collects the uploaded model parameters of all clients and performs global model update as follows:

$$\mathrm{E}\left(\mathbf{w}_{\mathrm{global}}\right) = \prod\nolimits_{k=1}^{N} \mathrm{E}\left(\mathbf{w}_{\mathrm{local},k}\right)^{\alpha_k} \ , \alpha_k = \left(\mathrm{D}_k\right) / \sum\nolimits_{i=1}^{N} \mathrm{D}_i \tag{3}$$

4) Decryption: Each clients decrypts E($\mathbf{w}_{\mathrm{global}}$) using the private key $sk(\lambda)$ by

$$\mathbf{w}_{\mathrm{global}} = \frac{L\left(\mathrm{E}\left(\mathbf{w}_{\mathrm{global}}\right)^{\lambda} \mathrm{mod} n^{2}\right)}{L\left(g^{\lambda} \mathrm{mod} n^{2}\right)} \bmod n \tag{4}$$



Original sample perturbations adversary sample

(**a**) image classification task

Original sample perturbations adversary sample

(**b**) fault diagnosis task

Figure 2: The visualization process of the adversary samples

4 Method

4.1 Overview

Fig. 3 presents the overview of the REFL, which mainly includes three stages: initialization, federated training, and model evaluation (Algorithm 1).

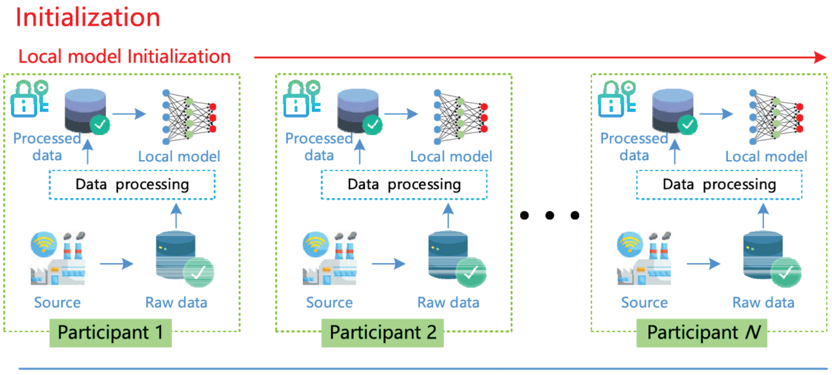


Figure 3: (Continued)

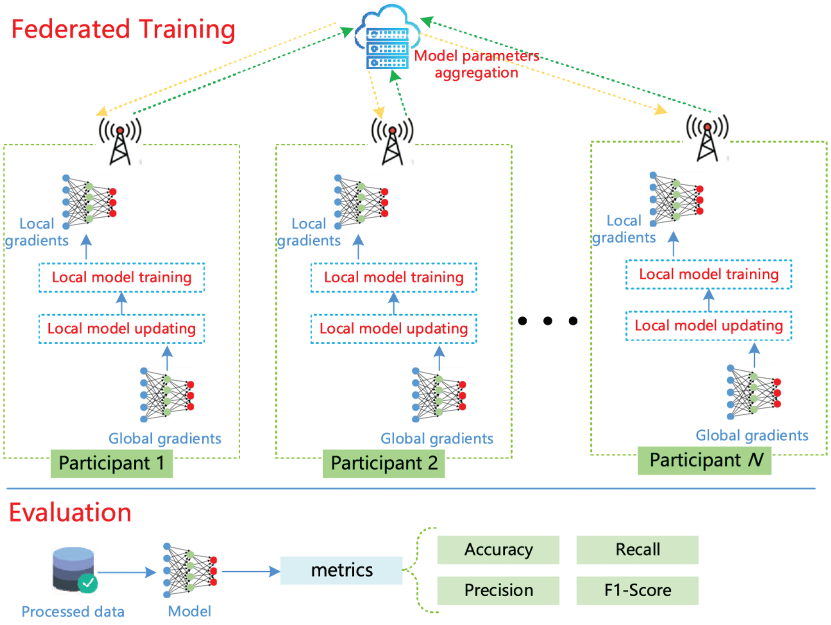


Figure 3: Overview of the REFL for machinery fault diagnosis

Algorithm 1: Robust and efficient federated learning

```
Input: $D_k$, training parameters, e.g., $\eta$, $b$, $\rho$ and Epoch.
Output: The global fault diagnosis model
Initialization:
a). Secure data transmission channels for the server and each
participant are established;
b). All the modules  of the system are initialized
Federated training:
For $e \leq Epoch$ do
    (I). Local training at the participants:
    for $k \leq N$ do //distributed training
        $\mathbf{x}_k, \mathbf{y}_k \leftarrow$ Dataloader($b$, $D_k$)//load training samples
        $\mathbf{x}'_k \leftarrow$ FSGM($\mathbf{x}_k, \mathbf{y}_k, \rho$)//generate adversary samples
        $\mathbf{g}_{\mathrm{local},k} \leftarrow$ SGD(model($\mathbf{w}_{\mathrm{local},k}, \mathbf{x}'_k$),$\mathbf{y}_k$)//calculate gradients
        $\mathcal{P}_K$ upload $\mathbf{g}_{\mathrm{local},k}$ to the server
    end
    (II). Parameter aggregation at the server:
    Receive $\mathbf{g}_{\mathrm{local},k}$ from the participants;
    Aggregate model parameters as shown in Eq. (5);
    Send $\mathbf{g}_{\mathrm{global}}$to all participants;
    (III). Local model updating at the participants:
    for $k \leq N$ do//distributed computing
        load the global model $\mathbf{g}_{\mathrm{global}}$
        $\mathcal{P}_K$ calculates the momentum terms
        $\mathcal{P}_K$ updates the local model weights
```

(Continued)

Algorithm 1 (continued)

```
            end
        end
    Evaluation:
    All participants use their local data to evaluate the model
    return Fault diagnosis model
```

1) Initialization stage: A representative industrial company is selected to initialize the intelligent model and training parameters, including the learning rate, and momentum rate. The initialized fault diagnosis model and training parameters are then broadcast to the other industrial participants.

(2) Federated training: In each training round, the local model gradients calculated by the participant are sent to the server, which then aggregates these parameters to update the global gradients. The participant downloads the updated global gradients from the cloud server, and updates the local model according to the optimization algorithm for the next local model training. Through multiple iterations with adversarial learning, the participants' local model can learn other participants' knowledge.

(3) Model evaluation: After finishing the federated training process, the participant evaluates the local fault diagnosis model on its testing dataset to confirm that the model reaches satisfactory performance.

4.2 Initialization Stage

First, the secure data transmission channels are established for the cloud server and the participants. Then, all modules of the REFL system are initialized. Recall that in this study, different participants use the same structure model and training parameters.

4.3 Local Model Training at the Participant

After initialization, local model training at the participant starts to train the fault diagnosis model with the gradient-based adversarial algorithm. First, each participant uses FSGM to generate adversarial samples with adversary rate ρ, which injects noise into the percentage ρ of the local training samples (detailed in Section 3.2). Then, SGD optimizer is used to calculate the gradients of the local fault diagnosis model. Thereafter, each participant uploads its model parameters $\mathbf{g}_{\text{local},k}$ to the cloud server.

4.4 Parameter Aggregation at the Server

The server recursively aggregates local gradients to derive the global gradients of the fault diagnosis model, and their expression is as follows:

$$\mathbf{g}_{\text{global}} = \sum_{k=1}^{N} \left(\alpha_k \cdot \mathbf{g}_{\text{local},k} \right) \tag{5}$$

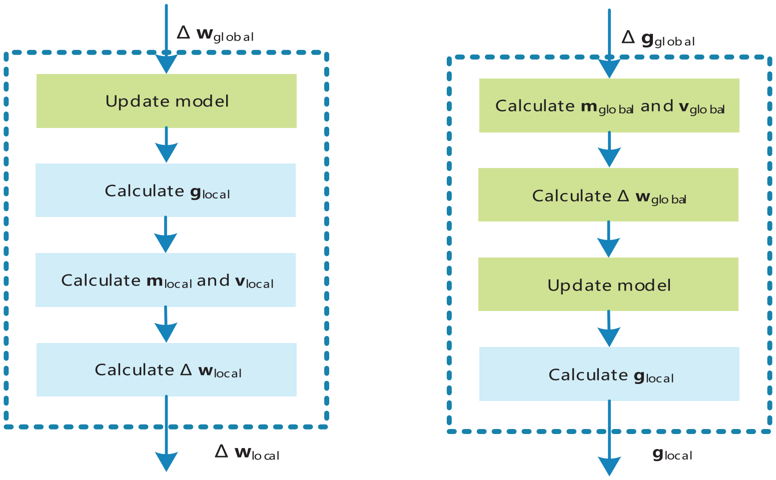
where $\mathbf{g}_{\text{global}}$ denotes the global gradients of the fault diagnosis model, and α_k denotes data contribution ratios calculated by $\alpha_k = b_k/b$.

4.5 Local Model Updating at the Participant

Upon receiving the global model parameters $\mathbf{g}_{\text{global}}$, the participant k loads the global model parameters to its local fault diagnosis model.

Much literature [31–33] has demonstrated that the Adam optimization algorithm is efficient when working with large problems involving a lot of model parameters. Intuitively, Adam hasthe advantages of the momentum gradient algorithm and the root mean square propagation algorithm. However, Sun et al. [34]

demonstrated that the Adam optimization algorithm is used directly by the participant in the FL system to update its local model, which would degrade the model performance, empirically in imbalance data distribution scenarios. In this study, to accelerate the model convergence speed and avoid performance degradation, we modify the local model updating process. Its core idea is that the participant uses the global gradients to update its model, not its local gradients, as shown in Fig. 4. This design not only essentially resolves the influence of unbalanced data distribution scenarios but also conforms to subsequent privacy preserving algorithms.



(**a**) our proposed scheme (**b**) traditional scheme

Figure 4: Comparison of our proposed scheme and the traditional scheme

The process of the local model training is expressed as follows:

(1) After receiving the global model gradients from the server, the participant computes momentum terms as

$$\mathbf{m}(t) = \beta_1 * \mathbf{m}(t-1) + (1-\beta_1) * \mathbf{g}_{\text{global}}(t) \tag{6}$$

$$\mathbf{v}(t) = \beta_2 * \mathbf{v}(t-1) + (1-\beta_2) * \mathbf{g}_{\text{global}}^2(t) \tag{7}$$

$$\hat{\mathbf{m}}(t) = \mathbf{m}(t) / \left(1-\beta_1^t\right) \tag{8}$$

$$\hat{\mathbf{v}}(t) = \mathbf{v}(t) / \left(1-\beta_2^t\right) \tag{9}$$

(2) The participant uses the above momentum terms to update the local model weights as

$$\mathbf{w}_{\text{local}} = \mathbf{w}_{\text{local}} - \eta \cdot \left(\hat{\mathbf{m}}(t) / \left(\sqrt{\hat{\mathbf{v}}(t)} + \varepsilon\right)\right) \tag{10}$$

Since the improved algorithm does not change the essence of the Adam algorithm, the convergence analysis of the algorithm in this paper is specifically introduced as follows [35]. Consider the objective function in the context of federated learning, where the goal is to minimize the expected risk. Assuming the objective function is convex loss with bounded gradients, $\|\nabla L(\theta_t)\|_2 \leq G$, $\|\nabla L_{\mathrm{t}}(\theta)\|_\infty \leq G_\infty$ for all θ, and distance between any θ_{t} generated by Adam is bounded. $\|\theta_{\mathrm{n}} - \theta_m\|_2 \leq D$, $\|\theta_{\mathrm{n}} - \theta_m\|_\infty \leq D_\infty$ for any m, $n \in \{1, ..., T\}$. We define $\gamma \triangleq \beta_1^2/\sqrt{\beta_2}$ and $\beta_{1,t} = \beta_1\lambda^{t-1}$. The values of parameters β_1 and β_2 fall within the range [0, 1], satisfying $\beta_1^2/\sqrt{\beta_2} < 1$.

It is worth noting that all participants have the same model weights since they use the same gradients to update their local models. Through the model updating process, the FL can accelerate the model convergence rate while avoiding performance loss. In addition, the local model updating scheme can be extended to other optimization algorithms according to task requirements.

5 Experimental Study

5.1 Fault Diagnosis Task

The Case Western Reserve University (CWRU) rolling bearing fault dataset has been utilized in multiple studies to evaluate the performance of diagnostic methods [36–39]. Hence, this paper also uses the fault dataset to evaluate the proposed REFL. As shown in Fig. 5, the experimental setup includes a motor, a torque transducer/encoder, a dynamometer, and auxiliary systems. The test bearings include three fault modes: ball fault, inner race fault and outer race fault. Each fault mode has three fault depths: 7, 14 and 21 mils. The information of CWRU fault dataset is listed in Table 2.

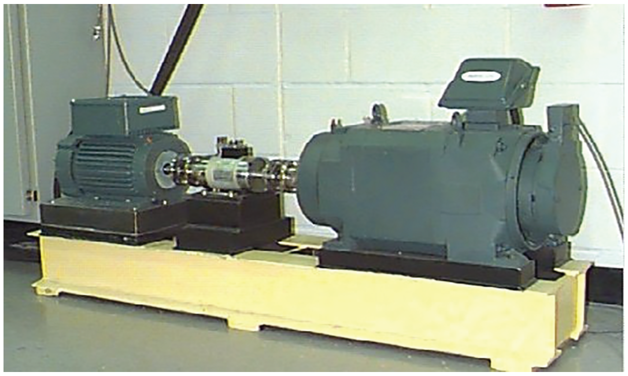


Figure 5: Experimental setup of CWRU fault diagnosis dataset

Table 2: CWRU fault dataset information

Label	Fault mode	Depth	Load (HP)
B007	Ball	7 mils	0, 1, 2
B014	Ball	14 mils	0, 1, 2
B021	Ball	21 mils	0, 1, 2
IR007	Inner race	7 mils	0, 1, 2
IR014	Inner race	14 mils	0, 1, 2
IR021	Inner race	21 mils	0, 1, 2
OR007	Outer race	7 mils	0, 1, 2
OR014	Outer race	14 mils	0, 1, 2
OR021	Outer race	21 mils	0, 1, 2
N0	Normal	0	0, 1, 2

Referring to these papers [21,40], the same data processing methods are adopted in this paper, whose detailed introduction is as follows:

(1) Data normalization: To enhance the convergence rate and precision, the raw data are normalized as

$$x_i = \left(x_i - \text{Mean}\left(\mathbf{x}\right)\right) / \sqrt{\text{Var}\left(\mathbf{x}\right)} \tag{11}$$

(2) Data segmentation: Instead of the fixed interval sliding window, the random sliding window is adopted to produce samples, as shown in Fig. 6.

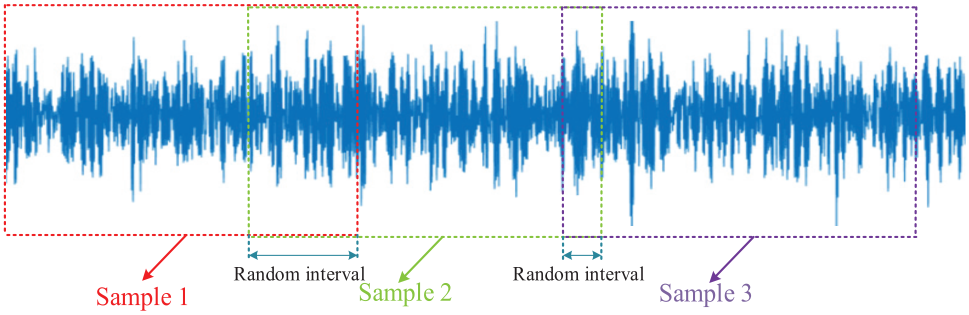


Figure 6: Random sliding window

(3) Data reconstruction: According to the input channels of the fault diagnosis model, one-dimension time-series samples are reconstructed into two-dimensional samples as presented in Fig. 7.

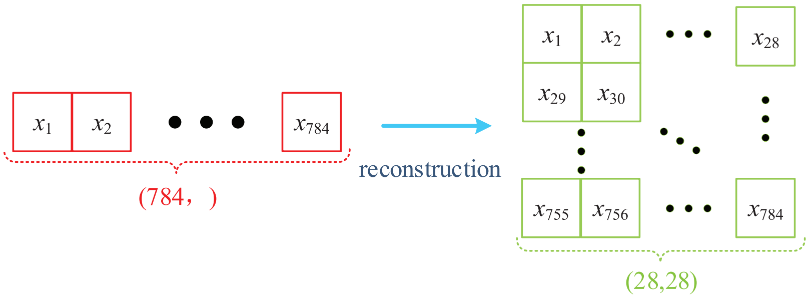


Figure 7: Data reconstruction process

Through the above preprocessing, the raw data are transformed into training samples for the fault diagnosis model. In real industrial scenarios, companies and factories often hold data in different modes. In this paper, the fault dataset is split into non-IID datasets for the FL system, which indicates that a participant holds only one mode of fault data. Therefore, the non-IID scenario is very suitable to verify the proposed fault detection method. The fault data is divided for the participants in the FL system.

5.2 Experimental Setting

1) Environment: Intel CPU i7-7700HQ, RAM 16 GB and NVIDIA 1070. The FL system is implemented by PyTorch 1.9.0 and CUDA 11.1.

2) Referring to prior studies, the specific deep learning model architecture is shown in Fig. 8.

3) We set epoch to 30, b to 32, η to 0.001, β_1 to 0.9, β_2 to 0.999, ρ to 0.4 and N to 5.

4) Metrics: Accuracy, precision, recall and F1-score serve as evaluation metrics to assess the diagnostic performance.

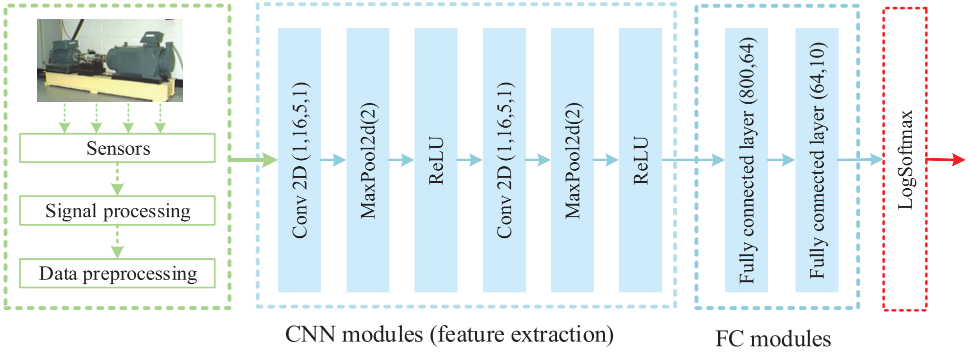


Figure 8: The architecture of the CNN-based fault detection model

5.3 *Case 1*

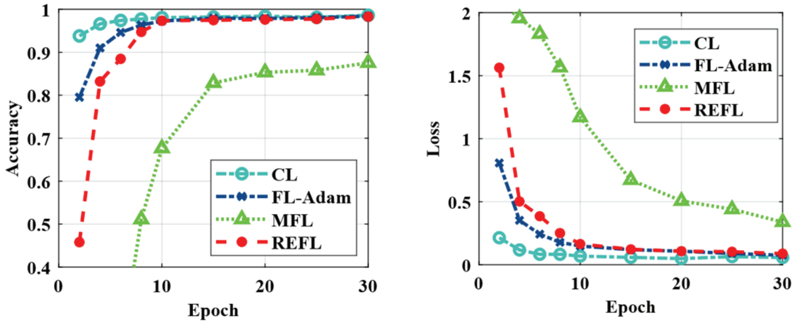
In this case, different learning approaches are implemented to demonstrate the superiority of the REFL, which are presented as follows:

1) CL: Centralized learning is used for comparisons, where aggregates all participants' local data for model training. Specifically, for the CL, the mini-batch size is set to 5 × 32.

2) MFL: Momentum federated learning uses the first-order momentum term to accelerate the model training process [27].

3) FL-Adam: In the FL-Adam system, the participant uses the Adam optimization algorithm to update their local model.

MFL and FL-Adam set the same training parameters as the REFL. Fig. 9 presents the training curves of different learning approaches, while Table 3 lists their diagnostic outcomes. From the accuracy and loss curves, Adam effectively accelerates convergence. Although the FL-Adam can improve training speed, its performance remains inferior to the CL and our proposed REFL. That is because the data distribution of the FL system is unbalanced. As Table 3 listed, REFL has better diagnostic competence compared with the FL-Adam and MFL. Through the above experiment results, it proves that the REFL can effectively improve the training speed while maintaining model accuracy.



(**a**) Accuracy (**b**) loss

Figure 9: Training curves of the different learning approaches

Table 3: Experiment results of different methods on CWRU fault dataset

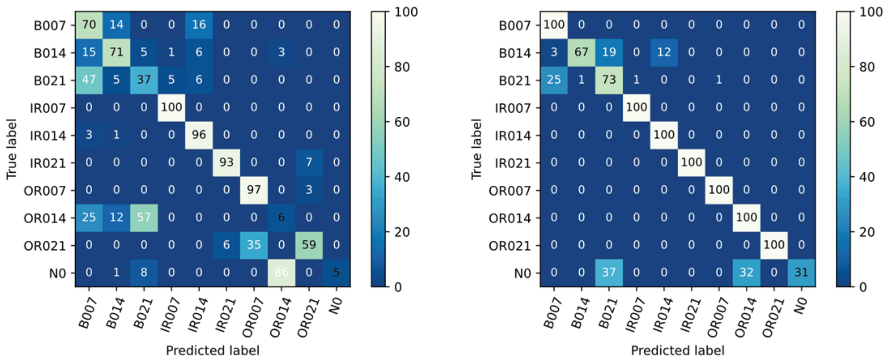
Method	Accuracy	Precision	Recall	F1-score
CL	98.87%	98.92%	98.87%	98.85%
FL-Adam	98.40%	98.46%	98.40%	98.37%
MFL	87.53%	87.82%	87.53%	86.08%
REFL	98.73%	98.78%	98.73%	98.71%

5.4 *Case 2*

In this case, to showcase the advantages of our proposed model, we add Gaussian noise with different levels to the testing datasets. The distributions of gaussian noises are set to (μ = 0, σ = 0.05), (μ = 0, σ = 0.1) and (μ = 0, σ = 0.2), respectively. The EFL approach consists of the FL and the proposed model updating scheme. SVM, CL and EFL are used to compare with the REFL. The diagnosis results of the different learning approaches on CWRU dataset with gaussian noise are listed in Table 4. Figs. 10 and 11 show the confusion matrixes and F1-Score of the different learning approaches on Testing-Gaussian (0, 0.2) dataset, respectively. It is evident that REFL significantly outperforms other methods. The fault diagnosis model trained by the REFL can effectively resist the negative effects of Gaussian noise, especially in Testing-Gaussian (0, 0.2) dataset. Therefore, our method can effectively resist external noise interferences by utilizing adversary samples.

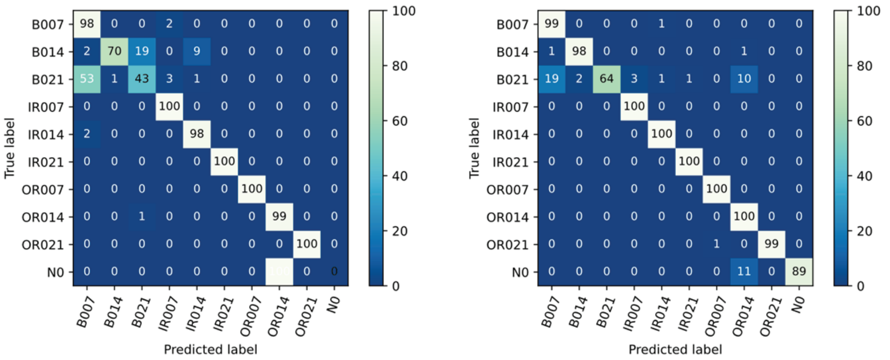
Table 4: Diagnosis results of different learning approaches on CWRU fault dataset with Gaussian noise

Method	Testing	Testing-Gaussian (0, 0.05)	Testing-Gaussian (0, 0.1)	Testing-Gaussian (0, 0.2)
SVM	82.23%	82.23%	82.20%	61.40%
CL	98.87%	98.87%	98.47%	87.00%
EFL	98.80%	98.67%	98.13%	80.73%
REFL	98.73%	98.67%	98.47%	95.00%



(**a**) SVM (**b**) CL

Figure 10: (Continued)



(**c**) EFL (**d**) REFL

Figure 10: Diagnostical matrixes of the different learning approaches on Testing-Gaussian (0, 0.2) dataset

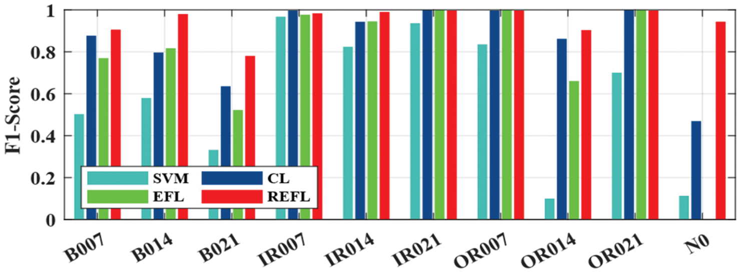
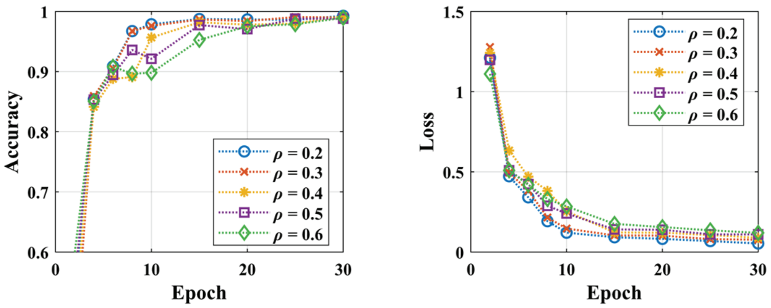


Figure 11: Different method F1-Score values on Testing-Gaussian (0, 0.2) dataset

5.5 *Case 3*

In this case, we assess the influence of the rate ρ on the REFL performance and present experimental results in Fig. 12 and Table 5. When the adversary range rate ρ rises from 0.2 to 0.6, the fault diagnosis performance of the REFL drops from 99.27% to 98.93%, but the resisting noise performance of the model is improved. Therefore, how to set ρ will be further investigated in the future.



(**a**) Accuracy (**b**) loss

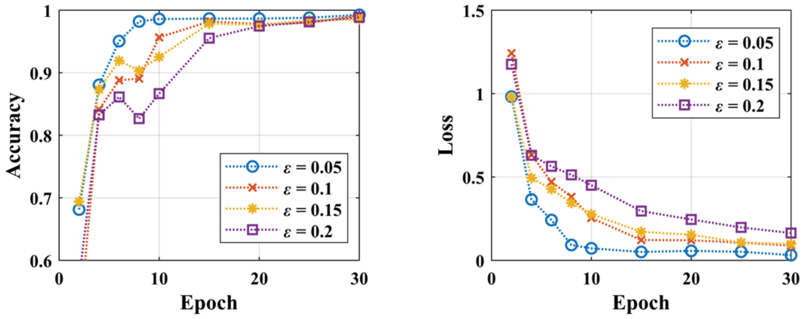
Figure 12: Training experiment curves of the REFL with different ρ

Table 5: Diagnosis results of the REFL with different ρ

Method	Testing	Testing-Gaussian (0, 0.05)	Testing-Gaussian (0, 0.1)	Testing-Gaussian (0, 0.2)
REFL (ρ = 0.2)	99.27%	99.27%	98.87%	87.13%
REFL (ρ = 0.3)	99.00%	98.93%	98.60%	88.60%
REFL (ρ = 0.4)	99.20%	99.13%	99.07%	88.93%
REFL (ρ = 0.5)	98.80%	98.73%	98.73%	91.00%
REFL (ρ = 0.6)	98.93%	98.73%	98.60%	94.80%

5.6 Case 4

In this case, we assess the influence of the rate ϵ on the REFL performance and present experimental results in Fig. 13 and Table 6. When ε rises from 0.05 to 0.2, the fault diagnosis performance of the REFL drops from 99.27% to 98.87%, but the resisting noise performance of the model is improved. Therefore, how to set ϵ will be further investigated in the future.



(**a**) Accuracy (**b**) loss

Figure 13: Training curves of the REFL with different ε

Table 6: Diagnosis results of the REFL with different ε

Method	Testing	Testing-Gaussian (0, 0.05)	Testing-Gaussian (0, 0.1)	Testing-Gaussian (0, 0.2)
REFL (ε = 0.05)	99.27%	99.27%	98.93%	86.27%
REFL (ε = 0.1)	99.20%	99.13%	99.07%	88.93%
REFL (ε = 0.15)	98.53%	98.53%	97.80%	92.40%
REFL (ε = 0.2)	98.87%	98.67%	97.47%	87.00%

6 Discussion

6.1 Comparison to Similar Studies

Traditional DL-based fault diagnosis methods require a lot of training samples in practical scenarios. To solve this issue, several recent similar studies also use federated learning to develop fault diagnosis methods with different participants. However, they do not consider protecting the data information of the industrial participants. Moreover, these prior works do not consider enhancing the robustness of the fault diagnosis methods. To improve the system efficiency, we modify the Adam algorithm for the FL paradigm to accelerate convergence speed while ensuring model accuracy under unbalance data scenarios. Considering the noise in the operation environment, the gradient-based adversary algorithm is introduced to boost the model robustness. Our proposed fault diagnosis method is verified through experimental cases in Section 5. The advantages of the REFL are as follows.

First, a federated learning framework is used to build the REFL for the machinery fault diagnosis. The structure of the REFL is inspired by this study. The Paillier encryption scheme is used to encrypt the gradients of the participants to preserve their local data information. Therefore, our proposed method can organize distributed participants to collaboratively build the fault diagnosis model.

Second, the REFL utilizes the first-order and second-order momentum terms to accelerate the convergence speed while ensuring model accuracy under unbalanced data scenarios, which can improve the operating costs of the FL system.

Third, the gradient-based adversary algorithm is adopted in the proposed method. The adversary algorithm can generate a lot of adversary samples at low resource costs. The model with adversary training has applicable diagnostic performance to resist the interference of noise.

6.2 Time of Training and Testing

An excellent fault diagnosis method requires accurate diagnosis and fast diagnosis speed. DL-based fault diagnosis methods require model training and then are applied to online diagnosis. Table 7 lists the training and testing times for REFL and SVM. By utilizing a GPU, the local model training time of the participant is faster than that of the SVM. However, the total consummation time of the REFL is still larger than that of SVM. As prior literature demonstrated, the training time does not need to be focused. The main concern is the testing time. From the table, our model testing time is approximately 0.20 s, which can be acceptable for online diagnostic tasks.

Table 7: Training and testing time of the REFL and SVM

Method	Training (s)	Testing (s)
REFL	13.63	0.20
SVM	18.92	8.79

7 Conclusions

This paper investigates federated learning methods in the field of machine fault diagnosis. Multiple companies can efficiently collaborate to build robust deep learning-based models while protecting their data resources. Considering the FL and industrial requirements, the operation steps of the Adam algorithm are modified to accelerate the model convergence speed and alleviate model performance degradation under unbalance data scenarios. Moreover, the gradient-based adversarial algorithm is first introduced to the fault diagnosis field, enhancing the model robustness against external noise. Four cases are conducted for validation, which demonstrates that the proposed method achieves outstanding performance. In the future, we plan to deploy and evaluate REFL in a truly distributed, multi-company setting to assess its performance and scalability within a live IoV ecosystem. Additionally, we will conduct systematic sensitivity analyses and develop dynamic adaptive strategies for the adversarial training parameters, aiming to automate the balance between model accuracy and robustness under varying operational conditions.

Acknowledgement: We acknowledge the support from all project team members and partner institutions.

Funding Statement: This work was supported in part by National key R&D projects (2024YFB4207203), National Natural Science Foundation of China (52401376), the Zhejiang Provincial Natural Science Foundation of China under Grant (No. LTGG24F030004), Hangzhou Key Scientific Research Plan Project (2024SZD1A24), "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2024C03254, 2023C03154), Jiangxi Provincial Gan-Po Elite Support Program (Major Academic and Technical Leaders Cultivation Project, 20243BCE51180).

Author Contributions: Zhen Wu: Methodology, Conceptualization, Model Construction; Hao Liu: Writing —Original Draft; Linlin Zhang: Writing—Review & Editing; Zehui Zhang: Software, Code Development, Experimental Validation; Jie Wu: Writing—Review & Editing; Haibin He: Writing—Review & Editing; Bin Zhou: Writing—Review & Editing. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The dataset analyzed in this study is the publicly available Case Western Reserve University (CWRU) Bearing Fault Dataset, which can be accessed at: https://engineering.case.edu/bearingdatacenter (accessed on 01 November 2025).

Ethics Approval: Not applicable. This article does not contain any studies with human participants or animals performed by any of the authors.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Min H, Fang Y, Wu X, Lei X, Chen S, Teixeira R, et al. A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis. Expert Syst Appl. 2023;224:120002. doi:10.1016/j.eswa.2023.120002.
2. Yan S, Sun W, Xia Y. A joint fault-tolerant and fault diagnosis strategy for multiple actuator faults of full-vehicle active suspension systems. IEEE Trans Autom Sci Eng. 2024;22:1928–40. doi:10.1109/TASE.2024.3372626.
3. Hossain MN, Rahman MM, Ramasamy D. Artificial intelligence-driven vehicle fault diagnosis to revolutionize automotive maintenance: a review. Comput Model Eng Sci. 2024;141(2):951–96. doi:10.32604/cmes.2024.056022.
4. Yu F, Chen G, Yang X, Gong Y, Huang Y, Du C. Engine misfire fault detection based on the channel attention convolutional model. Comput Mater Contin. 2025;82(1):843–62. doi:10.32604/cmc.2024.058051.
5. Che C, Wang H, Ni X, Fu Q. Domain adaptive deep belief network for rolling bearing fault diagnosis. Comput Ind Eng. 2020;143:106427. doi:10.1016/j.cie.2020.106427.
6. Wang X, Zhu H, Ning Z, Guo L, Zhang Y. Blockchain intelligence for Internet of vehicles: challenges and solutions. IEEE Commun Surv Tutor. 2023;25(4):2325–55. doi:10.1109/COMST.2023.3305312.
7. Khudhur AF, Kurnaz Türkben A, Kurnaz S. Design and develop function for research based application of intelligent Internet-of-vehicles model based on fog computing. Comput Mater Contin. 2024;81(3):3805–24. doi:10.32604/cmc.2024.056941.
8. Wu W, Joloudari JH, Jagatheesaperumal SK, Rajesh KNVPS, Gaftandzhieva S, Hussain S, et al. Deep transfer learning techniques in intrusion detection system-Internet of vehicles: a state-of-the-art review. Comput Mater Contin. 2024;80(2):2785–813. doi:10.32604/cmc.2024.053037.
9. Xu B, Li H, Ding R, Zhou F. Fault diagnosis in electric motors using multi-mode time series and ensemble transformers network. Sci Rep. 2025;15(1):7834. doi:10.1038/s41598-025-89695-6.
10. Ali Gultekin M, Bazzi A. Review of fault detection and diagnosis techniques for AC motor drives. Energies. 2023;16(15):5602. doi:10.3390/en16155602.
11. Jiang R, Ou S, Li B, Liu W, Cao B, Yu Y. A fault diagnosis method for typical failures of marine diesel engines based on multisource information fusion. Shock Vib. 2025;2025(1):1904885. doi:10.1155/vib/1904885.
12. Zhao Y, Song Z, Li D, Qian R, Lin S. Wind turbine gearbox fault diagnosis based on multi-sensor signals fusion. Prot Control Mod Power Syst. 2024;9(4):96–109. doi:10.23919/pcmp.2023.000241.
13. Wen J. A survey on federated learning: challenges and applications. Int J Mach Learn Cybern. 2023;14(2):513–35.

14. Liu W, Chen L, Chen Y, Zhang W. Accelerating federated learning via momentum gradient descent. IEEE Trans Parallel Distrib Syst. 2020;31(8):1754–66. doi:10.1109/TPDS.2020.2975189.
15. Zhang L, Zhang Z, Guan C. Accelerating privacy-preserving momentum federated learning for industrial cyber-physical systems. Complex Intell Syst. 2021;7(6):3289–301. doi:10.1007/s40747-021-00519-2.
16. Zhang W, Peng G, Li C, Chen Y, Zhang Z. A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals. Sensors. 2017;17(2):425. doi:10.3390/s17020425.
17. Hoang DT, Kang HJ. A survey on Deep Learning based bearing fault diagnosis. Neurocomputing. 2019;335(7):327–35. doi:10.1016/j.neucom.2018.06.078.
18. Khorram A, Khalooei M, Rezghi M. End-to-end CNN + LSTM deep learning approach for bearing fault diagnosis. Appl Intell. 2021;51(2):736–51. doi:10.1007/s10489-020-01859-1.
19. Kou L, Qin Y, Zhao X, Chen XA. A multi-dimension end-to-end CNN model for rotating devices fault diagnosis on high-speed train bogie. IEEE Trans Veh Technol. 2020;69(3):2513–24. doi:10.1109/TVT.2019.2955221.
20. Ben Abid F, Sallem M, Braham A. Robust interpretable deep learning for intelligent fault diagnosis of induction motors. IEEE Trans Instrum Meas. 2020;69(6):3506–15. doi:10.1109/TIM.2019.2932162.
21. Zhang Z, Xu X, Gong W, Chen Y, Gao H. Efficient federated convolutional neural network with information fusion for rolling bearing fault diagnosis. Control Eng Pract. 2021;116:104913. doi:10.1016/j.conengprac.2021.104913.
22. Li Y, Chen Y, Zhu K, Bai C, Zhang J. An effective federated learning verification strategy and its applications for fault diagnosis in industrial IoT systems. IEEE Internet Things J. 2022;9(18):16835–49. doi:10.1109/JIOT.2022.3153343.
23. Zhang W, Li X. Federated transfer learning for intelligent fault diagnostics using deep adversarial networks with data privacy. IEEE/ASME Trans Mechatron. 2022;27(1):430–9. doi:10.1109/TMECH.2021.3065522.
24. Ren Z, Zhu Y, Liu Z, Feng K. Few-shot GAN: improving the performance of intelligent fault diagnosis in severe data imbalance. IEEE Trans Instrum Meas. 2023;72:3516814. doi:10.1109/TIM.2023.3271746.
25. Wang B, Liang P, Zhang L, Wang X, Yuan X, Zhou Z. Enhancing robustness of cross-machine fault diagnosis via an improved domain adversarial neural network and self-adversarial training. Measurement. 2025;250:117113. doi:10.1016/j.measurement.2025.117113.
26. Wang X, Jiang H, Mu M, Dong Y. A trackable multi-domain collaborative generative adversarial network for rotating machinery fault diagnosis. Mech Syst Signal Process. 2025;224:111950. doi:10.1016/j.ymssp.2024.111950.
27. Wang D, Jin W, Wu Y, Ren J. Enhancing adversarial robustness for high-speed train bogie fault diagnosis based on adversarial training and residual perturbation inversion. IEEE Trans Ind Inform. 2024;20(5):7608–18. doi:10.1109/TII.2024.3363087.
28. Szegedy C. Intriguing properties of neural networks. arXiv:1312.6199. 2013.
29. Wang Y, Liu J, Chang X, Mišić J, Mišić VB. IWA: integrated gradient-based white-box attacks for fooling deep neural networks. Int J Intelligent Sys. 2022;37(7):4253–76. doi:10.1002/int.22720.
30. Goodfellow IJ, Shlens J, Szegedy C. Explaining and harnessing adversarial examples. arXiv:1412.6572. 2014.
31. Wang H, Li S, Song L, Cui L, Wang P. An enhanced intelligent diagnosis method based on multi-sensor image fusion via improved deep learning network. IEEE Trans Instrum Meas. 2019;69(6):2648–57. doi:10.1109/TIM.2019.2928346.
32. Gültekin Ö, Cinar E, Özkan K, Yazıcı A. Multisensory data fusion-based deep learning approach for fault diagnosis of an industrial autonomous transfer vehicle. Expert Syst Appl. 2022;200:117055. doi:10.1016/j.eswa.2022.117055.
33. Jais IKM, Ismail AR, Nisa SQ. Adam optimization algorithm for wide and deep neural network. Kno Eng Da Sc. 2019;2(1):41. doi:10.17977/um018v2i12019p41-46.
34. Sun H, Li S, Yu FR, Qi Q, Wang J, Liao J. Toward communication-efficient federated learning in the Internet of Things with edge computing. IEEE Internet Things J. 2020;7(11):11053–67. doi:10.1109/JIOT.2020.2994596.
35. Kinga D, Adam JB. A method for stochastic optimization. arXiv:1412.6980. 2014.
36. Yoo Y, Jo H, Ban SW. Lite and efficient deep learning model for bearing fault diagnosis using the CWRU dataset. Sensors. 2023;23(6):3157. doi:10.3390/s23063157.
37. Alonso-González M, Díaz VG, Pérez BL, G-Bustelo BCP, Anzola JP. Bearing fault diagnosis with envelope analysis and machine learning approaches using CWRU dataset. IEEE Access. 2023;11:57796–805. doi:10.1109/access.2023.3283466.

38. Raj KK, Kumar S, Kumar RR, Andriollo M. Enhanced fault detection in bearings using machine learning and raw accelerometer data: a case study using the case western reserve university dataset. Information. 2024;15(5):259. doi:10.3390/info15050259.
39. Banumalar K, Balakumar P. Efficient machine-learning model for bearing fault identification using the CWRU dataset. In: Control and information sciences. Singapore: Springer Nature; 2024. p. 489–507. doi:10.1007/978-981-97-5866-1_35.
40. Zhang Z, Guan C, Chen H, Yang X, Gong W, Yang A. Adaptive privacy-preserving federated learning for fault diagnosis in Internet of ships. IEEE Internet Things J. 2022;9(9):6844–54. doi:10.1109/JIOT.2021.3115817.