

Asynchronous Decentralized Federated Learning for Collaborative Fault Diagnosis of PV Stations

Qi Liu, Bo Yang, *Senior Member, IEEE*, Zhaojian Wang, *Member, IEEE*, Dafeng Zhu, Xinyi Wang, Kai Ma, *Member, IEEE*, and Xinpeng Guan, *Fellow, IEEE*

Abstract—Due to the different losses caused by various photovoltaic (PV) array faults, accurate diagnosis of fault types is becoming increasingly important. Compared with a single one, multiple PV stations collect sufficient fault samples, but their data is not allowed to be shared directly due to potential conflicts of interest. Therefore, federated learning can be exploited to train a collaborative fault diagnosis model. However, the modeling efficiency is seriously affected by the model update mechanism since each PV station has a different computing capability and amount of data. Moreover, for the safe and stable operation of the PV system, the robustness of collaborative modeling must be guaranteed rather than simply being processed on a central server. To address these challenges, a novel asynchronous decentralized federated learning (ADFL) framework is proposed. Each PV station not only trains its local model but also participates in collaborative fault diagnosis by exchanging model parameters to improve the generalization without losing accuracy. The global model is aggregated distributedly to avoid central node failure. By designing the asynchronous update scheme, the communication overhead and training time are greatly reduced. Both the experiments and numerical simulations are carried out to verify the effectiveness of the proposed method.

Index Terms—Federated learning, asynchronous decentralized learning, collaborative fault diagnosis, multiple PV stations.

1 INTRODUCTION

THE large consumption of fossil fuels has caused serious issues, such as energy shortages, environmental pollution, and global warming [1]. As a renewable and sustainable energy source, solar energy is attracting more and more attention, where photovoltaics is one of the primary ways [2]. The photovoltaic (PV) arrays are the most fundamental part of the PV system and are prone to suffer from various faults due to the harsh outdoor environment [3]. Since these faults have different levels of harm to the PV system, it is of great significance to accurately diagnose the types of faults, thereby reducing the economic loss of power generation and avoiding safety accidents.

In the last decades, many studies focused on PV fault diagnosis have been carried out. The existing methods can be mainly categorized into three types: sensing-based methods [4], [5], [6], electrical characteristics-based methods [7], [8], [9] and machine learning-based (ML) methods [10], [11]. The sensing-based methods and electrical characteristics-based methods rely on manual fault feature extraction and some key parameters need to be determined through specific experiments according to the topology, which reduces the generalization of models and is not suitable for increasingly complex PV stations. Compared with them, the ML-based methods are data-driven and automatically extract fault

features through a large number of samples, which has occupied a dominant position in the field of fault diagnosis.

However, the existing ML-based methods developed fault diagnosis models based on a single PV station without considering the case that both the types and number of fault samples are insufficient. They either assumed that the collected fault samples were sufficient and improved the accuracy of fault diagnosis through advanced neural networks, or assumed that the number of fault samples was insufficient but already contained all types of faults to be diagnosed. Although these methods are practically important and achieve high accuracy, it is equally significant to study scenarios involving multiple PV stations to collaboratively train a shared fault diagnosis model, which is considered an important way to break data island [12], [13], [14]. In fact, it is difficult for a single PV station to collect samples of all fault types, and there is an unbalanced distribution of samples from multiple PV stations. If these samples can be fully utilized together, the generalization of the model can be significantly improved without losing accuracy. Thus, it is important to design a new ML framework to make multiple PV stations learn the fault features contained in the local data from each other.

To design a new collaborative fault diagnosis framework for multiple PV stations, the following three challenges need to be addressed: data island, framework robustness, and modeling efficiency. First, the collected samples contain the privacy information of the PV system. Since multiple PV stations belong to different operators, they are not allowed to share the original data directly. A privacy-preserving data interaction framework should be adopted. Second, the validity and accuracy of the fault diagnosis model are

- Qi Liu, Bo Yang, Zhaojian Wang, Dafeng Zhu, Xinyi Wang, and Xinpeng Guan are with the Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China; Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai 200240, China; Shanghai Engineering Research Center of Intelligent Control and Management, Shanghai 200240, China. Kai Ma is with School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China. Bo Yang is the corresponding author. E-mail: bo.yang@sjtu.edu.cn

essential to the safe and stable operation of the PV system. The robustness of the collaborative fault diagnosis modeling must be guaranteed. Third, the modeling efficiency is seriously affected by model update strategy due to different amounts of data and computing capabilities [15]. Hence it is important to design a novel global model aggregation scheme to reduce the communication overhead and training time.

In this paper, we propose a novel federated learning (FL) method that combines the decentralized (serverless) framework and asynchronous update strategy for PV fault diagnosis. The problem of the unbalanced distribution of fault samples and data island among multiple PV stations is addressed through collaborative fault diagnosis while ensuring data privacy, as well as the framework robustness and modeling efficiency issues. The proposed method fully utilizes the data of multiple PV stations to improve the generalization of the model without losing accuracy and is suitable for various meteorological conditions.

We mainly focus on three common PV array faults in this paper: 1) short-circuit faults, defined as the accidental connection or low impedance between two points in the PV array [16]; 2) degradation faults, defined as the increase in the equivalent series resistance or the decrease in the parallel resistance after a certain time of operation [17]; 3) partial shading faults, defined as the reduction of actual effective irradiance of the PV module due to surface dust, clouds, or other objects blocking the sunlight [18].

Our major contributions are summarized as follows:

- This paper focuses on the collaborative modeling method in PV fault diagnosis, which enables multiple PV stations to learn the fault features from others without sharing original data, and improves model generalization. To solve the problem of server paralysis in the existing centralized FL methods, a novel asynchronous decentralized federated learning (ADFL) framework is proposed to achieve the distributed aggregation of global models.
- Considering the difference in computing capability and amount of data for each PV station, a new global model aggregation algorithm is proposed to dynamically determine the participating agents in each round. An asynchronous update scheme is designed to accelerate the global model aggregation, which greatly reduces communication overhead and training time. The theoretical convergence analysis of the proposed algorithm is given.
- Extensive experiments are carried out based on simulation and real-world data to show that our method enables each PV station to accurately identify more fault types and is suitable for various meteorological conditions.

The remainder of this paper is organized as follows. Section 2 reviews the related work. Section 3 describes the system model. Section 4 introduces the ADFL based collaborative fault diagnosis method. Section 5 presents the experimental and numerical simulation results. Section 6 concludes the paper.

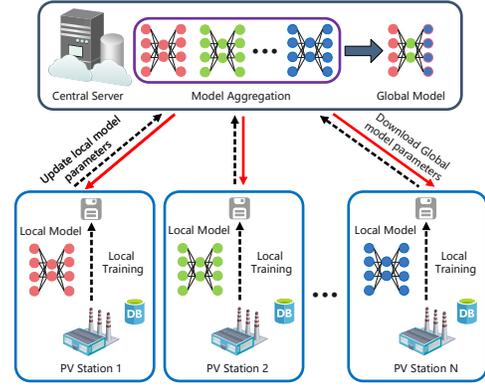


Fig. 1. The traditional centralized FL framework

2 RELATED WORK

In recent years, ML technology has been widely used in PV fault diagnosis, such as deep neural networks (DNNs) [19], support vector machines (SVM) [11], [20] and extreme learning machine (ELM) [21], [22], etc. However, these methods rely on a large amount of labeled PV data, which is usually difficult or expensive to obtain. Aiming at the problem of insufficient fault samples, Feng *et al.* [10] integrated domain knowledge into the learning process and proposed a domain-knowledge-based deep-broad learning framework (DK-DBLF) to reduce demand for labeled samples and improve the model flexibility. Zhao *et al.* [23] proposed a graph-based semi-supervised learning model, which only used a few labeled training data to detect line-line faults and open-circuit faults in PV arrays. Lu *et al.* [24] proposed a deep convolutional generative adversarial network (DA-DCGAN) to convert normal data from the target domain into virtual fault data by employing domain adaptation. However, the conventional ML methods mentioned above do not consider the case that both the types and number of fault samples are insufficient. They are mostly based on the assumption that samples of all fault types to be diagnosed have been included in the labeled dataset.

FL is a privacy-preserving distributed machine learning paradigm [25]. It is suitable for scenarios where the distribution of fault samples of multiple PV stations is unbalanced and they cannot share the original data directly due to privacy. Specifically, as illustrated by Fig. 1, participants in FL only share model parameters without providing their original data. After that, a shared global model is obtained by aggregating the model parameters of all participants, which has excellent accuracy and generalization. The concept of FL was introduced by Google [26] and has been used in several fields [27], [28], [29], [30]. Y. Liu *et al.* proposed an FL-based gated recurrent unit neural network for traffic flow prediction [27]. Y. Gao *et al.* in [28] proposed a federated region-learning (FRL) framework to comprehensively use the information of multiple weather monitoring stations for weather forecasting. S. Lee *et al.* in [29] applied federated reinforcement learning to home energy management systems to reduce energy costs. W. Zhang *et al.* proposed an FL method based on dynamic fusion to collaboratively detect COVID-19 infection [30]. However, the methods mentioned above either used centralized FL or updated the global

model synchronously. Once the central server crashes, the collaborative modeling cannot be carried out. In addition, since the difference in computing capability and amount of data is not considered, long waiting time for synchronous update mechanism results in low modeling efficiency.

3 SYSTEM MODEL

In this section, we first introduce the system configuration and then present the proposed ADFL framework.

3.1 System Configuration

We assume that there are N PV stations participating in collaborative fault diagnosis, and they belong to different owners. All of these PV stations have data storage, processing, and transmission capability, and use their data to train the local fault diagnosis models based on the neural networks. We regard each PV station as an agent, and use the set $\mathcal{C} = \{C_1, C_2, \dots, C_N\}$ to represent all agents, where C_i is responsible for PV station i . As shown in Fig. 2, each agent is composed of several arrays and has its local dataset D_i . Specifically, the I-V characteristic curves, corresponding temperature and irradiance of each array can be collected by the I-V tester equipped with an environmental tester.

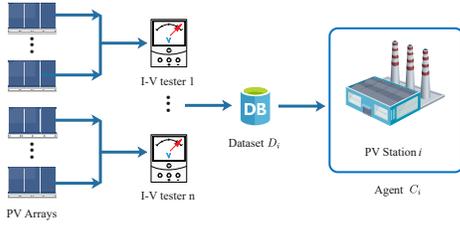


Fig. 2. The structure of each agent

Considering the situation where both the types and number of samples are insufficient, we assume that each agent only has part of fault samples, and the fault types involved in the local data of each agent are not the same. To implement the global model aggregation, each agent is supposed to use neural networks with the same structure to train its local fault diagnosis model.

3.2 ADFL Framework

Different from the centralized FL methods, the proposed ADFL framework is fully decentralized and does not require the participation of the central server, as shown in Fig. 3. Each agent first records the I-V characteristic curves and environmental information as the dataset to train the local fault diagnosis model, and then broadcasts the model parameters to other participants while receiving updated models sent by them. More details are described as follows:

3.2.1 Functions of agents

Each agent mainly contains four functions: (a) Update the local fault diagnosis model based on its fault samples; (b) Distribute local model parameters to other agents; (c) Receive updated models from other agents and complete

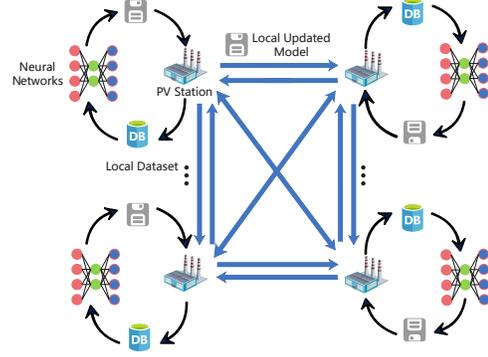


Fig. 3. The proposed ADFL framework

model aggregation distributedly; (d) Compare the aggregated global model with the current local model, and reserve the model with better accuracy as the new local model for subsequent communication. Repeat the above steps until the global model converges. Since all agents in the FL framework use the neural networks with the same structure, the global model is obtained by aggregating the model parameters of multiple PV stations without sharing the original data.

In the process of collaborative fault diagnosis modeling, each agent is both a client and a server, as shown in Fig. 4. The model receiving termination threshold is introduced as a flag to update the global model. Each agent checks the current model receiving queue after updating its local model. Once the termination threshold is triggered, it only sends model parameters to the agents in the queue to complete the global model update. Meanwhile, it sends skip signals to the remaining agents and clears the queue, indicating that they skip the current round of model aggregation. The data volume of the skip signal is very small and is not considered to consume communication resources. Furthermore, an optimal model selection mechanism is executed to compare the latest global model with the current local model. The model with higher fault diagnosis accuracy is retained for the next round of communication.

3.2.2 Asynchronous update strategy

Due to the difference in computing capability and the amount of data, the time required for each agent to update the local model is different. The synchronous update scheme leads to low modeling efficiency due to the long waiting time. This motivates us to design an asynchronous strategy, where each agent joins the global model aggregation in different rounds, as shown in Fig. 5.

Since the proposed ADFL framework is decentralized, the global model is aggregated without considering the status of other agents when the model parameters collected by any agent reach a certain amount. In the synchronous update strategy, a certain number of agents are reselected in each round to train their models, and the slowest agent greatly affects the update efficiency [31]. Although the number of agents participating in each round is certain due to the termination threshold, the proposed asynchronous strategy performs more efficiently. This is because not all of these agents start to update their local models from the round

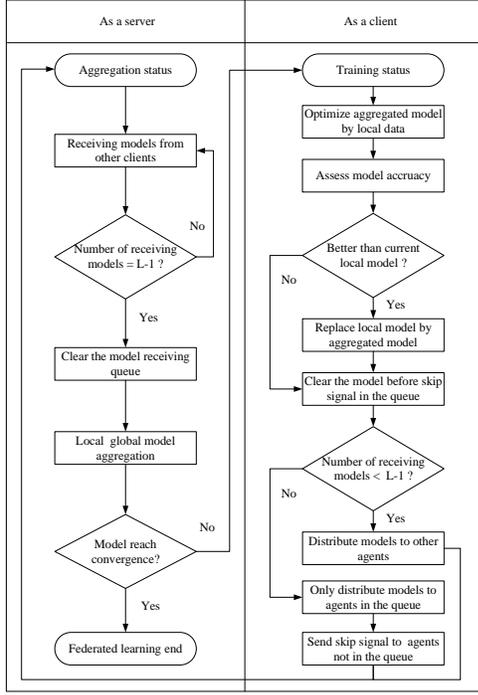


Fig. 4. Two roles of each agent

in which they participate in global model aggregation. For agents with low computing capability, it takes a long time to update their local model parameters while the global model may have been updated several times by other agents during this time. When the local model update is completed, they will immediately participate in the global model aggregation. For example, agent C joins the global update in round 2, but it starts to train the local model since round 1, which effectively reduces the waiting time. It should be stressed that even if agent A crashes in round 3, agent B and agent C still update the global model. Compared to the existing centralized FL framework, the proposed ADFL method has several advantages.

- *Save computing resources*: It enables the local model updated by each agent to be used in a specific global model aggregation round rather than being discarded directly like the synchronous update mechanism, which avoids wasting local computing resources of each participant.
- *Enhance model diversity*: It ensures that agents with different computing capabilities and amounts of data have the chance to join the model aggregation, which enhances the diversity and generalization of the model.
- *Improve training efficiency*: It speeds up the aggregation frequency of the global model and saves time waiting for other agents to update, thereby improving training efficiency.
- *Guarantee the robustness*: The global model is aggregated distributedly, which avoids the issue that collaborative fault diagnosis cannot be carried out once the central node crashes to a certain extent.

4 ADFL FOR COLLABORATIVE FAULT DIAGNOSIS

In this section, we specifically introduce the implementation of the collaborative fault diagnosis based on the proposed ADFL framework. It is mainly composed of two parts, including each agent separately training the local fault diagnosis model and multiple agents jointly aggregating a shared global model.

4.1 Data Preprocessing

The I-V characteristic curves of PV arrays under different operating status are quite different [32]. Specifically, the I-V characteristic curves of the normal and three fault states under the temperature of 20°C and the irradiance of $600\text{W}/\text{m}^2$ are shown in Fig. 6.

The original I-V curves are not directly used for fault diagnosis for two reasons. First, there are many sampling points for the collected I-V characteristic curves, which contain redundant information and occupy additional computing resources. Second, the sampling points on the curves are unevenly distributed, which makes it difficult to accurately reflect the fault features. To solve these problems, the original I-V characteristic curves are compressed and enhanced by combining the down-sampling and bilinear interpolation methods. To retain as much effective feature information as possible, 20 new data points V_{sample_i} and I_{sample_i} are sampled equidistantly in the range of $[0, V_{\text{oc}}]$ and $[0, I_{\text{sc}}]$ respectively, $i = 1, 2, \dots, 20$. The corresponding current I_n and voltage values V_n , $n = 1, 2, \dots, 20$, are calculated from the original data points according to bilinear interpolation. Then, they are sorted in ascending order of voltage. Specifically, the voltage and current of the 40 new sampling points are respectively calculated by (1) and (2):

$$I_n = \frac{(V_{\text{sample}_i} - V_{\text{left}}) \cdot I_{\text{right}} + (V_{\text{right}} - V_{\text{sample}_i}) \cdot I_{\text{left}}}{V_{\text{right}} - V_{\text{left}}}, \quad (1)$$

$$V_n = \frac{(I_{\text{sample}_i} - I_{\text{left}}) \cdot V_{\text{right}} + (I_{\text{right}} - I_{\text{sample}_i}) \cdot V_{\text{left}}}{I_{\text{right}} - I_{\text{left}}}, \quad (2)$$

where V_{left} , I_{left} and V_{right} , I_{right} respectively represent the current and voltage values of the left and right samples closest to V_{sample_i} and I_{sample_i} .

Specifically, the I-V characteristic curves of partial shading faults before and after data preprocessing under the temperature of 20°C and the irradiance of $600\text{W}/\text{m}^2$ are shown in Fig. 7. Through the above data preprocessing method, the original curve is compressed from approximately 400 data points to 40, which are evenly distributed on the curve. The current and voltage of the new I-V curve are expressed as a 40×2 I-V vector. Meanwhile, considering the influence of temperature and irradiance, we add them to the I-V vector as a 40×2 environmental feature vector. Then the reconstructed 40×4 two-dimensional array is regarded as a fault sample and used as the input of the later CNN.

4.2 CNN Unit

In the proposed ADFL framework, the global model is obtained by exchanging and aggregating local model parameters between agents. Therefore, a neural network with the same structure is required to train a local fault

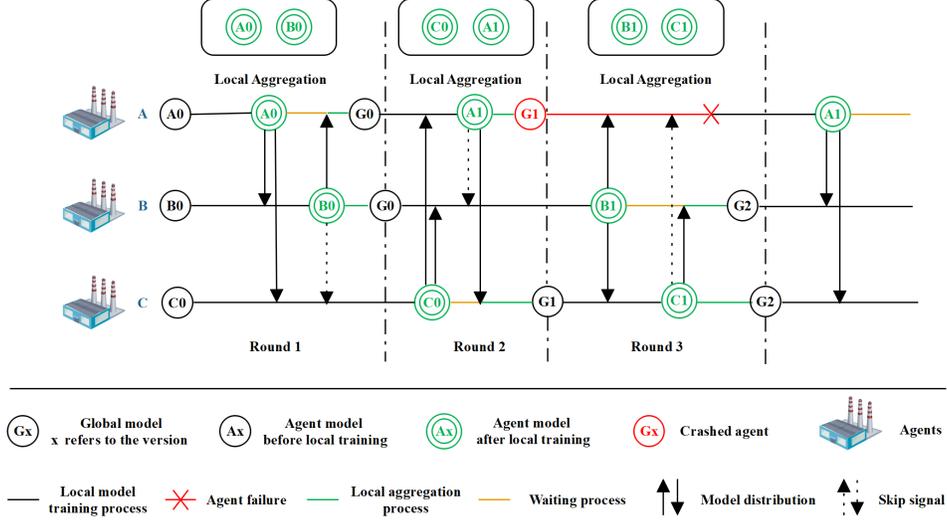


Fig. 5. The diagram of asynchronous update strategy

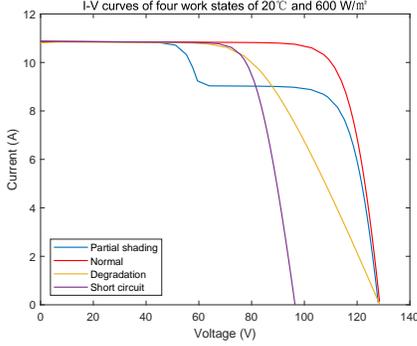


Fig. 6. I-V characteristic curves under different work status

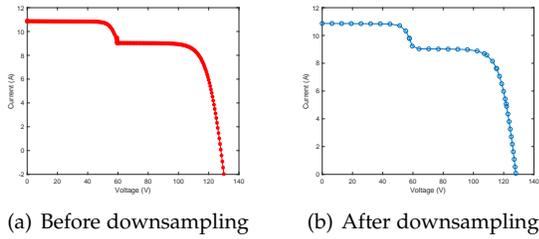


Fig. 7. I-V characteristic curve of partial shading fault

diagnosis model for each agent separately. Considering that the collected samples are 40×4 feature matrices, a plain CNN proposed in [33] is selected to automatically extract the two-dimensional data features and train the local fault diagnosis model. The structure of the CNN includes a 2-D CNN layer, 1-D CNN layers, Maxpool layers, a fully connected (FC) layer, and the Softmax function. The architecture of the network is shown in Table. 1. The following subsection specifically introduces the process of collaborative fault diagnosis modeling for multiple PV stations.

TABLE 1
Configuration for the plain CNN

Layer	Output dimension	Detailed architecture of CNN
2-D CNN	$37 \times 1 \times 1$	$k = 44, C_{out} = 1, ss = 1, p = 0$
Squeeze	37×1	Dimension squeeze
1-D CNN	18×3	$k = 3, C_{out} = 3, ss = 2, p = 0$
1-D CNN	18×5	$k = 3, C_{out} = 5, ss = 1, p = 1$
MaxPool 1D	8×5	$k = 4, ss = 2$
1-D CNN	8×8	$k = 3, C_{out} = 8, ss = 1, p = 1$
MaxPool 1D	4×8	$k = 2, ss = 2$
1-D CNN	4×16	$k = 3, C_{out} = 16, ss = 1, p = 1$
MaxPool 1D	1×16	$k = 4, ss = 1$
FC	10	Fully-connected Layer
Linear Classifier	10×9	FC + Softmax

4.3 Collaborative Fault Diagnosis Modeling

The goal of collaborative fault diagnosis is to enable multiple PV stations to identify more fault types by making full use of their local data while ensuring data privacy. Therefore, an asynchronous decentralized aggregation algorithm is proposed to obtain a high-precision collaborative (global) model by integrating their local model parameters. Specifically, based on the local data set D_k and the current local model w_t^k , agent C_k obtains the updated local model w_{t+1}^k by running the adaptive moment estimation (Adam) [34] method on a mini-batch of $B < D_k$, and thus we have:

$$f(x^{(i)}; w) = \hat{y}_i, \quad (3)$$

$$J(w) = \frac{1}{B} \sum_{i=1}^B [y_i \log \hat{y}_i + (1 - y_i) \log (1 - \hat{y}_i)], \quad (4)$$

$$g_t = \frac{1}{B} \nabla_w \sum_{i=1}^B J(\hat{y}_i, y^{(i)}), \quad (5)$$

$$m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t, \quad (6)$$

$$v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2, \quad (7)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1 t}, \hat{v}_t = \frac{v_t}{1 - \beta_2 t}, \quad (8)$$

$$w_{t+1}^k = w_t^k - l_r \frac{\hat{m}_t}{\sqrt{v_t + \epsilon}}, \quad (9)$$

where $x^{(i)}$ is a sample in a small batch B and the model parameters of the CNN are represented by w . \hat{y}_i is the probability vector of the CNN softmax function and y_i is the real label of the sample. B is the mini-batch size and g represents the gradient of the loss function J . The initial values of m_t and v_t are equal to 0 when $t = 0$. β_1 and β_2 are the average moving coefficients. l_r is the initial learning rate and ϵ is a small constant to avoid a zero denominator. In the following, we take agent C_k as an example to illustrate.

During each round of communication, agent C_k sends its local model parameters w_t^k to other agents, and receives weights w_t^i from others. Based on the collected parameters, each agent aggregates global models distributedly. Once receiving the model parameters of other participants, it immediately updates the global model without waiting for other agents to complete the update. To further improve the communication efficiency, we set a termination threshold L for asynchronous updates. L is a positive integer between 1 and C , and C is the total number of agents. By properly modifying the value of the termination threshold L based on specific problems, the ideas in this paper can be applied to the collaborative training of various numbers of agents. \mathcal{L} is a set containing the agents whose model parameters have been received. After an agent receives model weights from other $L - 1$ participants, it starts to aggregate global model G_t . Thus we have:

$$G_t = \sum_{i \in \mathcal{L}} \alpha_i w_t^i + \sum_{j \notin \mathcal{L}} \beta_j w_{t-1}^j, \quad (10)$$

where α_i and β_j are the mixing weights for the models. The global model of consists of two parts. For the first L participants (including itself), the agent obtains their latest local model weights w_t^i , and for the remaining participants, the expired local weights w_{t-1}^j are used instead.

The selection of mixing weights is significant to the effectiveness of model aggregation. It is worth noting that the training data of each participating agent is changing in each round. Therefore, data size is significant to model aggregation. Inspired by [35], the following mixing weights method is chosen:

$$\alpha_i = \frac{d_i}{D}, \beta_j = \frac{d_j}{D}, \quad (11)$$

where $d_i = |D_i|$ and $d_j = |D_j|$ denote the local data size of agent C_i and agent C_j , respectively. The data size of all agents is represented as $D = \sum |D_i|$.

To speed up the convergence, the optimal model selection mechanism is also designed. When each agent participates in global model aggregation for the first time, the aggregated model G_t is directly regarded as the new local model w_{t+1}^k . Otherwise, agent C_k compares the accuracy of the latest global model G_t and the current local model w_t^k . The model with higher accuracy is selected as the local fault diagnosis model w_{t+1}^k for subsequent rounds of global model aggregation. The detailed steps are shown in Algorithm 1.

Algorithm 1 Asynchronous Decentralized Model Aggregation

Require:

- Local datasets D_k ;
- Local model w_t^k ;
- Model aggregation round t ;
- Mini-batch size B
- Latest local weights from $L - 1$ agents w_{t+1}^i ;
- Expired local weights from other agents w_t^j ;
- Initialization CNN parameters $m_0, v_0, l_r, \epsilon, \beta_1$ and β_2

Ensure:

- Best local fault diagnosis model w_{t+1}^k
 - 1: **Local Model Learning:** For agent C_k , parameters update is performed to train the local fault diagnosis model as $w_{t+1}^k = w_t^k - l_r \frac{\hat{m}_t}{\sqrt{v_t + \epsilon}}$ by running Adam method over a number of mini-batches.
 - 2: **Model Broadcasting:** Agent C_k broadcasts its local model w_t^k to other participants;
 - 3: **Asynchronous Decentralized Model Aggregation:** Once agent C_k receives the model parameters of other participant C_i , it immediately calculates the global model components $\alpha_i w_t^i$. After receiving the model parameters of $L - 1$ other agents, the global model is aggregated as $G_t = \sum_{i \in \mathcal{L}} \alpha_i w_t^i + \sum_{j \notin \mathcal{L}} \beta_j w_{t-1}^j$;
 - 4: **Local Model Optimization:** For the first time the agent C_k participates in the global model aggregation, $w_{t+1}^k = G_t$. Otherwise, agent C_k compares the fault diagnosis accuracy of the aggregated global model G_t and the current local model w_t^k . Then, the model with higher accuracy is selected as new local model w_{t+1}^k ;
 - 5: **return** w_{t+1}^k ;
-

4.4 Convergence Analysis

The convergence of the global model of FL has been proved by some studies. However, since we design a new decentralized FL framework and adopt an asynchronous model aggregation mechanism, it is necessary to analyze the convergence of Algorithm 1.

In the proposed ADFL method, the following optimization problems need to be solved in each round of global model update:

$$\arg \min_G F(G, t) = \sum_{i \in \mathcal{L}} \frac{d_i}{D} F_i(G) + \sum_{j \notin \mathcal{L}} \frac{d_j}{D} F_j(G), \quad (12)$$

where G represents the global model parameters, and $F_k(G)$ is the average loss of the local data on agent C_k . $F_k(G)$ is calculated by (13):

$$F_k(G) = \frac{1}{d_k} \sum_{(x_k, y_k) \in D_k} f(G; x_k, y_k), \quad (13)$$

where $f(\cdot)$ denotes the loss function, and $d_k = |D_k|$ represents the local data size of agent C_k .

We prove the convergence of the proposed ADFL method by analyzing the upper bound of $F(G(t), t) - F(G^*; t)$, where G^* denotes the optimal parameters of the global model. We mainly refer to the idea of convergence analysis in [36] and [37], and make appropriate changes based on our situation. First, we also make the following assumption:

Assumption 1. For each agent C_k , the loss function $F_k(w)$ is convex, ρ -Lipschitz and H -smooth. For any w, w' , we have:

- 1) $F(w) - F(w') \leq \nabla F(w)^T (w - w')$,
- 2) $\|F_k(w) - F_k(w')\| \leq \rho \|w - w'\|$,
- 3) $\|\nabla F_k(w) - \nabla F_k(w')\| \leq H \|w - w'\|$.

Based on Assumption 1 and triangle inequality, we further introduce the definition of gradient divergence in [36], to quantify the gradient difference between the local and global loss function. As defined in [36], the δ_k is the upper bound of the gradient divergence of agent C_k , and $\bar{\delta}$ denotes the upper limit of δ_k :

$$\|\nabla F_k(w) - \nabla F(w)\| \leq \delta_k \leq \bar{\delta}. \quad (14)$$

Considering that each agent trains the local model τ epochs before sharing parameters, we use e to represent the index of the epoch, $e = 1, 2, \dots, \tau$. Formally, we define a virtual global model $G_t([e])$, which is obtained by the aggregation of all $w_i^k([e])$ at epoch e . Inspired by [37], we introduce an auxiliary model $z_t([e])$, which is initialized as G_{t-1} in round t and updated by centralized gradient descent for optimizing the same target $F(G, t)$. Based on the definition, we have:

$$G_t([e]) = \sum_{i \in \mathcal{L}} \frac{d_i}{D} w_i^i([e]) + \sum_{j \notin \mathcal{L}} \frac{d_j}{D} w_j^j([e]), \quad (15)$$

$$w_i^i([e]) = w_i^i([e-1]) - l_r \nabla F_i(w_i^i([e-1])), \quad (16)$$

$$w_j^j([e]) = w_j^j([e-1]), \quad (17)$$

$$z_t([e]) = z_t([e-1]) - l_r \nabla F(z_t([e-1])), \quad (18)$$

where $w_i^i([e])$ is updated from $w_i^i([e-1])$ and $w_j^j([e])$ does not change with epoch e because agent j does not participate in the t -th round of model aggregation.

Theorem 1: For any epoch e in round t , we have the loss divergence bound:

$$F(G_t([e])) - F(z_t([e])) \leq \rho \bar{u}(e), \quad (19)$$

where

$$\bar{u}(x) \triangleq \frac{\bar{\delta}}{H} ((l_r H + 1)^x - 1) - l_r \bar{\delta} x. \quad (20)$$

Proof. The proof of Theorem 1 is mainly based on Lemma 2 in [36] and the details are in Appendix A.

Theorem 1 quantifies the difference in loss between the $G_t([e])$ aggregated by the local models and the $z_t([e])$ learned from centralized training in each epoch $[e]$. It is worth noting that $G_t([e]) = z_t([e]) \leq \rho \bar{u}(\tau)$ at $e = \tau$ and $\tau = 1$ since $\bar{u}(1) = 0$. Considering that $G_t = G_t([t \cdot \tau])$, the convergence upper bound of G_t is further analyzed in Theorem 2.

Theorem 2: The convergence upper bound of the global model in round t with τ epochs is given by:

$$F(G_t) - F(G^*) \leq \frac{1}{T(\chi\varphi - \zeta)}, \quad (21)$$

when it satisfies the following conditions:

- 1) $l_r \leq \frac{1}{H}$,
- 2) $\chi\varphi - \zeta > 0$,
- 3) $F(z_t([e])) - F(G^*) \geq \varepsilon, \forall e$,
- 4) $F(G_t) - F(G^*) \geq \varepsilon$,

where $\varepsilon > 0$, $T = t\tau$, $\chi \triangleq \min_t \frac{1}{\|z_t([0]) - G^*\|^2}$, $\varphi \triangleq l_r(1 - \frac{Hl_r}{2})$, and $\zeta \triangleq \frac{\rho \bar{u}(\tau)}{\tau \varepsilon^2}$.

Proof. Condition 1) indicates that the learning rate l_r cannot be too large. Condition 2) restricts the auxiliary model $z_t([e])$ from deviating too much from the virtual global model $G_t([e])$. Conditions 3) and 4) ensure that the loss of the approximate model is not less than the theoretical optimal model G^* [37]. By combining the result of Theorem 1 and the Lemmas in [36] and [38], Theorem 2 is proved and see Appendix B for details.

5 EXPERIMENTS AND SIMULATION STUDIES

In this section, experiments and numerical simulations are carried out to verify the effectiveness of the proposed ADFL method. First, experiments based on the temperature and irradiance data in a small range are carried out to illustrate the basic properties of the method. Then, numerical simulation is investigated to verify that the method is suitable for various meteorological conditions.

5.1 Experiments Configuration

In the experiments, we consider the collaborative fault diagnosis of three agents. Each agent has a certain number of local fault samples, which are divided into the training set and test set at a ratio of 7 : 3. It is worth noting that each fault type is divided according to this ratio instead of randomly shuffled samples. This ensures that all the local fault types are covered in the training set, which can maximize the accuracy and generalization performance of the local model. The model receiving termination threshold L is set to two. Some key hyperparameter settings are shown in Table 2. The PV module installed in the actual PV array is HT60-156M-C-330, and the detailed parameters are shown in Table 3.

TABLE 2
Settings of the hyper-parameters

Hyper-parameters	Value	Hyper-parameters	Value
B	128	l_r	1e-3
β_1	0.995	β_2	0.999
Epoch	50	ϵ	1e-8

TABLE 3
Technical data at standard test condition

Module parameters	Value
Maximum power point M_{pp} (W)	99.925
Open circuit voltage V_{oc} (V)	21.5
Short-circuit current I_{sc} (A)	6.03
Maximum power voltage V_{mp} (V)	17.5
Maximum power current I_{mp} (A)	5.71
Fuse current (A)	15
Maximum system voltage (V)	1000

The actual PV array consists of two parallel-connected PV strings with twenty-two PV modules in series. The I-V tester (Model: PROVA1011) equipped with an environmental tester with Bluetooth communication function is selected to collect data in various situations, as shown in Fig. 8. Based on the actual PV array, short-circuit faults, partial shading faults, and degradation faults are simulated respectively. Specifically, the short-circuit faults are realized by connecting the two panels directly with wires, the partial shading

TABLE 4
Dataset configuration for each agent in six real experiments

Sample numbers of each experiment	Types	Normal	Short-current	Degradation	Partial shading
Participants					
Agent 1		100/ 100/ 100/ 100/ 100/ 100	100/ 100/ 100/ 100/ 100/ 100	0/ 100/ 0/ 100/ 100/ 100	0/ 0/ 100/ 100/ 0/ 100
Agent 2		100/ 100/ 100/ 100/ 100/ 100	0/ 0/ 0/ 0/ 0/ 100	100/ 100/ 100/ 100/ 100/ 100	0/ 0/ 0/ 0/ 100/ 100
Agent 3		100/ 100/ 100/ 100/ 100/ 100	0/ 0/ 0/ 0/ 100/ 100	0/ 0/ 0/ 0/ 0/ 100	100/ 100/ 100/ 100/ 100/ 100

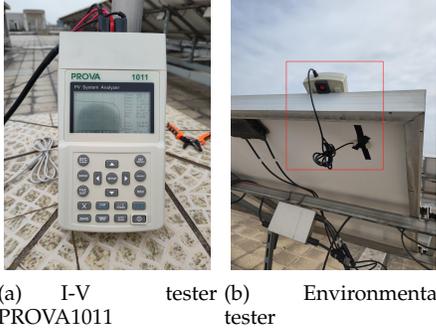


Fig. 8. The I-V tester PROVA1011 and environmental tester

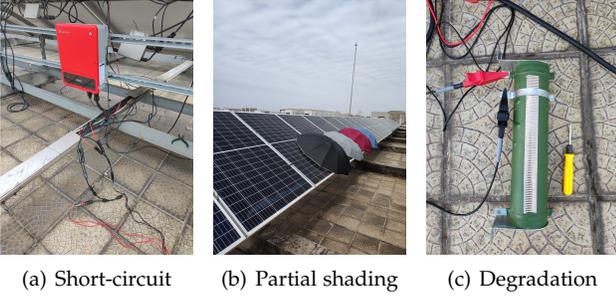


Fig. 9. Fault simulation in actual PV array

faults are implemented by covering part of the PV panels with umbrellas, and the degradation faults are simulated by connecting a high-power resistor of 8 ohms in series to the PV array, as shown in Fig. 9. The I-V characteristic curves and corresponding temperature and irradiance of normal and three fault states are collected under different weather conditions. For each state, 100 pieces of data are collected and reconstructed into 40×4 two-dimensional matrices as diagnostic samples through the data preprocessing method described above.

The experiments are implemented by using Pytorch and are conducted on HP ZBook Create G7 with Intel (R) Core (TM) i9-10885H CPU, 32GB RAM, NVIDIA GeForce RTX 2070 Max-Q.

5.2 Experimental Results

To verify the effectiveness of the proposed method in practical applications, six groups of experiments are carried out based on the actual PV array in terms of accuracy, communication overhead, and training time, as shown in Table 4. In each experiment, we set the local data of different agents to contain different types of fault samples. Specifically, if the local data of an agent contains a certain type, the number of samples is 100. Otherwise, it is 0. Taking agent

1 as an example, the first row below the normal column, i.e. **100 / 100 / 100 / 100 / 100**, indicates that the sample numbers in the normal state of agent 1 are 100 in all six experiments. These samples are further divided into the training set and test set by 7:3 as described in Section 5.1. To ensure the validity of the accuracy comparison, centralized FL and centralized training use the same dataset of the above six experiments as ADFL. It is worth noting that only centralized training can share data. At this time, the training sets of the three agents are combined to train the model in each experiment.

5.2.1 Accuracy of Local Model without ADFL

We first evaluate the local and global fault diagnosis accuracy of the local models of the three agents in the six groups of experiments without introducing the proposed ADFL method, as shown in Fig. 10. It is worth noting that the local here refers to the samples owned by the agent itself, and the global refers to all the samples of the three agents.

Obviously, in all experiments, each agent reaches 99% accuracy on its local dataset, which shows that the CNN used for training accurately extracts fault features. However, it is difficult for a single PV station to effectively diagnose all the fault types based on its local model in the case that both the amount and types of fault samples are insufficient. Specifically, taking experiment 4 as an example, the global fault classification confusion matrix of the three agents is plotted in Fig. 11. It can visually reveal the accuracy of the classifier for each type of fault sample. It can be seen that the global fault diagnosis accuracy of agent 2 and agent 3 is only 62.5% because the CNN cannot learn the fault characteristic of the missing types. Only when the agent has samples of all fault types, the local model can achieve high global accuracy, such as agent 1. Therefore, it is significant to achieve collaborative fault diagnosis of multiple PV stations through the proposed ADFL method, which improves the generalization ability of the model.

5.2.2 Accuracy of The Global Model with ADFL

Collaborative fault diagnosis modeling is carried out under six groups of experiments to verify the effectiveness of the ADFL method. The global accuracy of the collaborative models varies with aggregation rounds and is further compared with the centralized FL and the centralized training, as shown in Fig. 12. Since the centralized training does not involve the model aggregation and has nothing to do with the rounds, its accuracy is represented by a straight line with a constant value.

It can be observed that the proposed method converges to 99% accuracy after a certain round of aggregation, which is almost the same as the centralized training method. Taking agent 2 as an example, the global accuracy in the

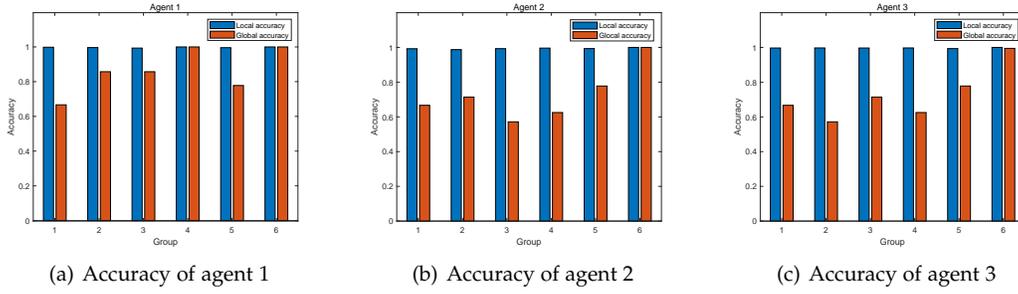


Fig. 10. The fault diagnosis accuracy of each agent without the proposed method

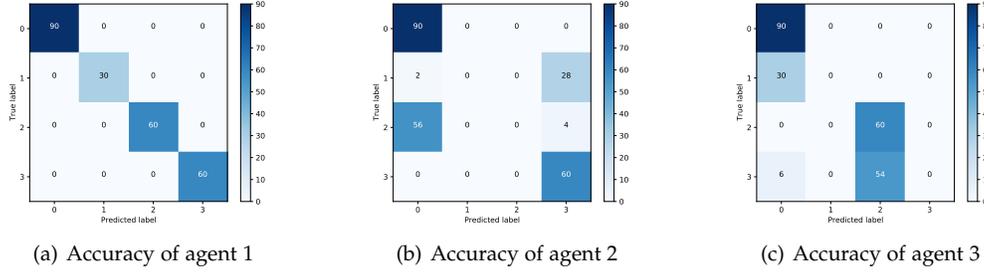


Fig. 11. The global fault classification confusion matrix of the three agents

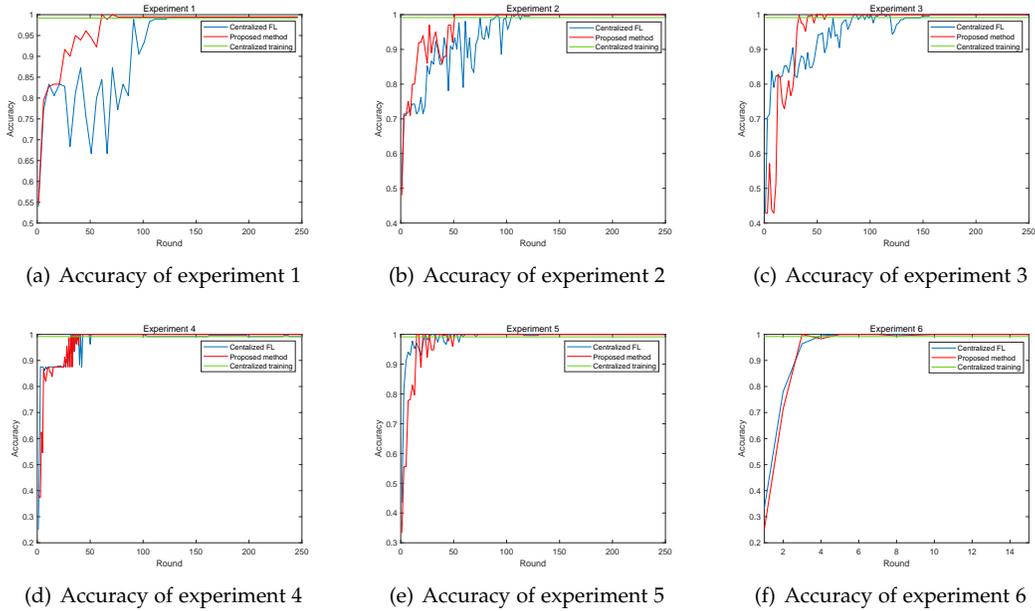


Fig. 12. Global fault diagnosis accuracy of six real experiments

first five groups of experiments is improved by about 33%, 28%, 43%, 38%, and 23% respectively. This is mainly because the proposed ADFL method enables the collaborative model to learn the fault features in the local data of all agents while ensuring data privacy. In addition, compared with the traditional centralized FL, the proposed method achieves convergence with fewer aggregation rounds.

5.2.3 Communication and Training Efficiency

The main factors affecting communication overhead include the number of parameters transmission in each round

of global model aggregation and the number of rounds required for model convergence. In general, the greater the local sample difference between agents, the more aggregation rounds required for the global model to converge. In this paper, the communication overhead is set as the total number of parameters transmission when the global model converges. The communication overhead is compared with that of centralized FL under six groups of experiments, as shown in Fig. 13 (a). Obviously, the proposed method significantly reduces the communication overhead in the

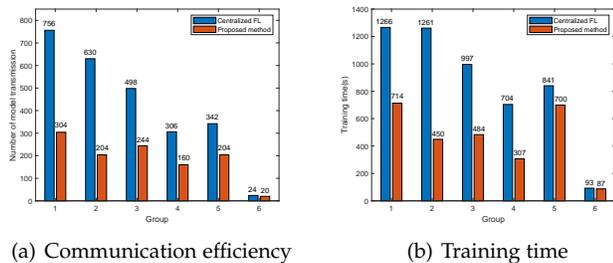


Fig. 13. Communication overhead and training time in experiments

process of model aggregation, especially when the local data of agents is greatly heterogeneous. For example, the communication overhead is reduced by approximately 60%, 67%, and 51% in experiment 1, experiment 2, and experiment 3. Even if the fault samples of agents are similar, this method still saves more than 16% of the transmission costs, such as experiment 6.

Although the above results have shown that the proposed method converges faster to high accuracy and reduces the communication overhead, this may not necessarily reduce the time of collaborative fault diagnosis modeling, because the time consumed in each round is also a key factor [39]. Especially for power systems, the speed of data processing is particularly important [40], [41]. To further evaluate the training efficiency of the proposed framework, the training time required for global model convergence in six groups of experiments is recorded separately and compared with that of centralized FL, as shown in Fig. 13 (b). In each round of local updates, the epochs of local model training are set to 50 for agents. It can be seen that the proposed method greatly reduces the model training time, especially when the samples vary greatly between agents. For example, the training time is reduced by approximately 65% in experiment 2.

5.3 Numerical Simulation Results

Due to the limitation of weather and environmental conditions, the actual experimental samples are collected in situations where the temperature and irradiance change within a small range. To further verify that the proposed method is suitable for various meteorological conditions, a wider range of samples need to be collected. Therefore, a PV module fault simulation model based on Simulink is built as shown in Fig. 14. The simulation model consists of three parallel-connected PV strings with six PV modules in series. To save space, only three PV panels are shown in each column in the figure, where the second and third panels actually represent two and three PV modules respectively.

Specifically, short-circuit faults are simulated by connecting a resistor R_{short} with a small resistance in parallel to the PV module. A resistor $R_{\text{degradation}}$ is connected in series on the trunk road of the PV array to increase the equivalent series resistance to simulate the degradation faults. Partial shading faults are simulated by connecting a gain module $Gain_{\text{PS}}$ with a value between 0 and 1 in series between the PV panels and the irradiance module. Some parameter settings of PV modules in the simulation model are listed in Table 6.

By setting the temperature module and irradiance module of the simulation model, the temperature change range is set from 10°C to 70°C , and the step length is 2°C . At the same time, the irradiance change range is set from $50\text{W}/\text{m}^2$ to $1000\text{W}/\text{m}^2$, and the step length is $10\text{W}/\text{m}^2$ [33]. The I-V characteristic curves and the corresponding temperature and irradiance under all of these situations are collected by the I-V tester and temperature irradiance recording module. Based on the simulation model, we select samples in the status of normal, short-circuit, degradation, and partial shading. There are 2976 samples of each state and the total is 11904. We also design six groups of numerical simulation experiments to evaluate the effectiveness of the proposed ADFL method in terms of accuracy, communication overhead, and training time, as shown in Table 5. The structure and hyperparameter settings of CNN are the same as the previous experiments.

5.3.1 Accuracy of The Global Model with ADFL

The six experiments designed in Table 5 can be divided into three cases. Experiment 1 represents the first case, where the local data of each agent contains normal data and a kind of fault data, as shown in Fig. 15(a). Moreover, the fault type of each agent is different, which has the strongest data heterogeneity. Experiments 2, 3, 4, and 5 represent the second case, where the local dataset of each agent contains normal data and a part of fault data, as shown in Fig. 15(b)(c)(d)(e). The fault types of various agents overlap to a certain extent, which reduces the heterogeneity of data. Experiment 6 represents the third case, where the local data of each agent contains normal data and all three types of fault data, as shown in Fig. 15(f). At this time, the fault types of each agent are the same, which conforms to independent and identical distribution (IID), and the data heterogeneity is the lowest. To avoid accidental errors, the result of each experiment is the average accuracy of 20 times, as shown in Fig. 15.

It can be observed that the proposed method is significantly better than the centralized FL in convergence speed for the first case. For the second case, the proposed method still converges slightly faster than centralized FL under various data heterogeneity situations. For the third case, the convergence speed of the proposed method is almost the same as that of the centralized FL due to low data heterogeneity and few rounds required for convergence. In all cases, the proposed method can achieve 99% accuracy after certain aggregation rounds. In fact, there is significant data heterogeneity among actual PV stations due to differences in geographic environment and operating conditions, and the proposed method is more suitable to be adopted.

5.3.2 Communication and Training Efficiency

Under six groups of simulation experiments, the communication overhead and training time of the proposed method are compared with the centralized FL respectively, as shown in Fig. 16. The proposed method converges with lower communication overhead and less training time, which is consistent with previous experimental conclusions.

The above experimental and numerical simulation results show that the proposed ADFL method is applicable to

TABLE 5
Dataset configuration for each agent in six numerical simulation experiments

Sample numbers of each simulation	Types	Normal	Short-current	Degradation	Partial shading
Participants					
Agent 1		2976/ 2976/ 2976/ 2976/ 2976/ 2976	2976/ 2976/ 2976/ 2976/ 2976/ 2976	0/ 2976/ 0/ 2976/ 2976/ 2976	0/ 0/ 2976/ 2976/ 0/ 2976
Agent 2		2976/ 2976/ 2976/ 2976/ 2976/ 2976	0/ 0/ 0/ 0/ 0/ 2976	2976/ 2976/ 2976/ 2976/ 2976/ 2976	0/ 0/ 0/ 0/ 2976/ 2976
Agent 3		2976/ 2976/ 2976/ 2976/ 2976/ 2976	0/ 0/ 0/ 0/ 2976/ 2976	0/ 0/ 0/ 0/ 0/ 2976	2976/ 2976/ 2976/ 2976/ 2976/ 2976

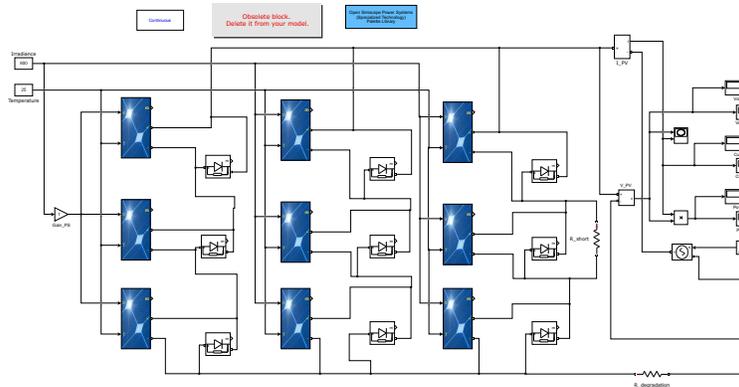


Fig. 14. PV module fault simulation model

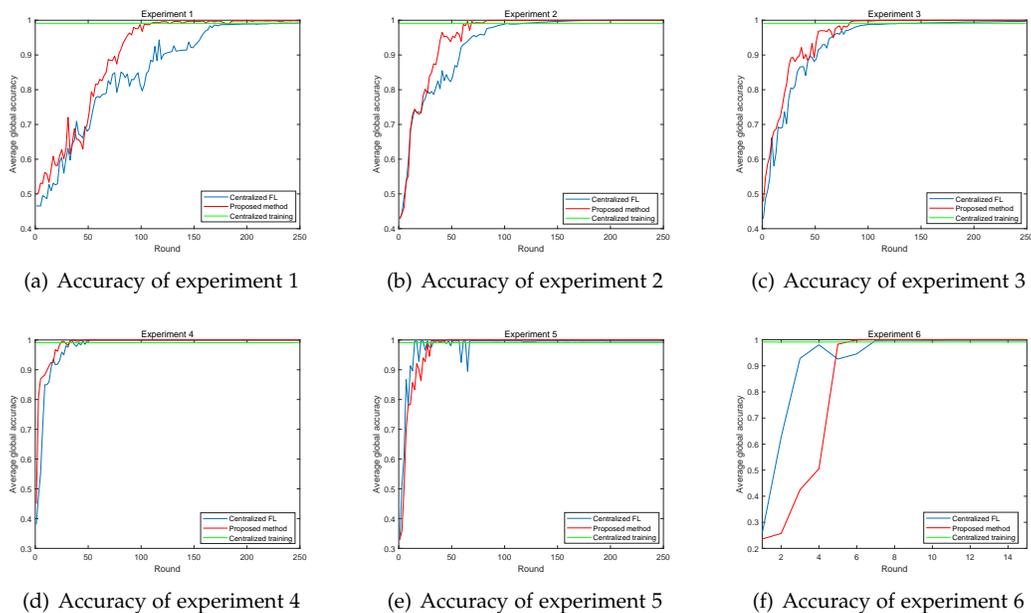


Fig. 15. Average accuracy of six simulation experiments

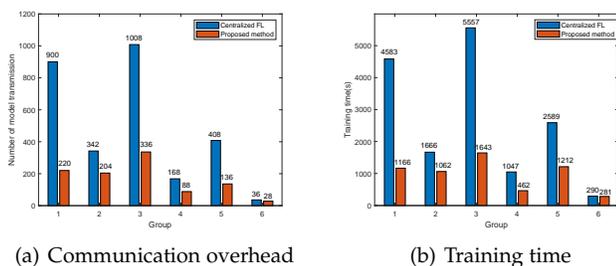


Fig. 16. Communication overhead and training time in simulations

various meteorological conditions. Through the collaborative fault diagnosis of multiple PV stations, the problem of uneven distribution of fault samples is well solved.

5.3.3 Optimality analysis

To further prove that the proposed method will not converge to the local optimum in a single dataset, for the six simulation experiments in Section 5.3.1, we test the average accuracy of the global model 20 times, as shown in Fig. 17. The error bar clearly shows the average accuracy and standard deviation of each round. Obviously, the optimal model selection mechanism performs better in terms of average accuracy, convergence speed and fluctuation, which

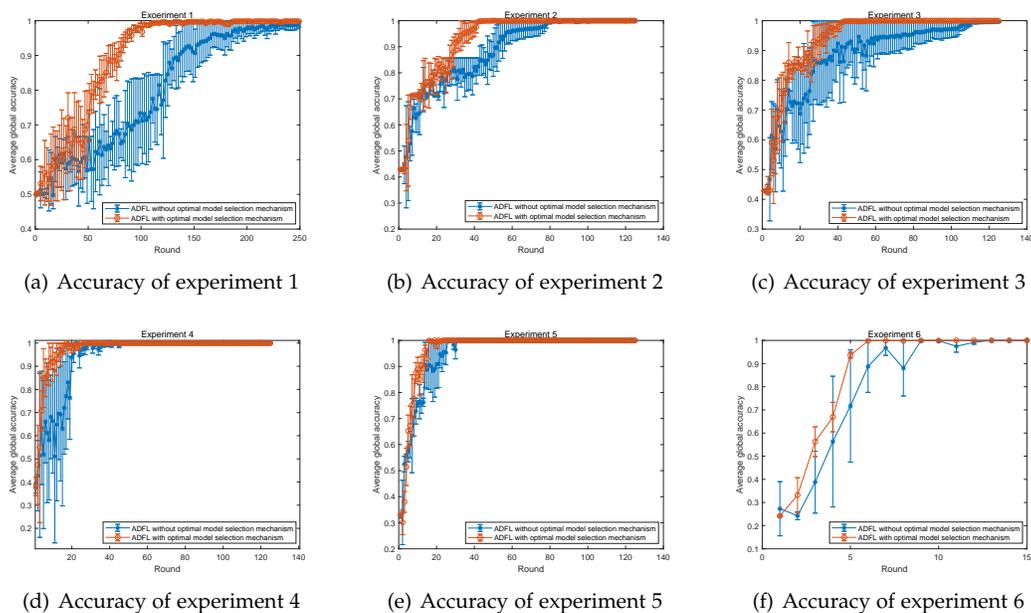


Fig. 17. Average accuracy of six simulation experiments

TABLE 6
Setting of the simulation module parameters

Module parameters	Value
Short-circuit resistance R_{short} (ohms)	0.001
Degradation resistance $R_{degradation}$ (ohms)	3
Partial shading gain $Gain_{PS}$	0.5
Maximum power point M_{pp} (W)	99.925
Open circuit voltage V_{oc} (V)	21.5
Voltage at maximum power point V_{mp} (V)	17.5
Short-circuit current I_{sc} (A)	6.03
Current at maximum power point I_{mp} (A)	5.71
Light-generated current I_L (A)	6.0576
Diode saturation current I_0 (A)	2.0517e-10
Diode ideality factor	0.96445
Shunt resistance R_{sh} (ohms)	551.8793
Series resistance R_s (ohms)	0.2392
Cells per module (Ncell)	36
Temperature coefficient of V_{oc} (%/deg.C)	-0.36
Temperature coefficient of I_{sc} (%/deg.C)	0.06

TABLE 7
The longest continuous upload times of the same model for agents

Participants	Experiments					
	11	2	3	4	5	6
Agent 1	5	6	6	3	2	0
Agent 2	6	9	6	6	4	1
Agent 3	26	15	13	12	17	2

indicates that it is sufficient to select the model by comparing the accuracy.

In the meanwhile, the average longest continuous upload times of the same model for each agent with the optimal model selection mechanism are recorded, as shown in Table 7. It is worth noting that the values in Table 7 are rounded up. It can be seen that the number of consecutive uploads of the same model for each agent is not high, which also indicates that the proposed method will not converge

to the local optimum in a single dataset.

6 CONCLUSION

In this paper, a novel asynchronous decentralized federated learning framework is proposed for PV fault diagnosis. The data island problem among multiple PV stations is addressed through collaborative fault diagnosis, which enhances the generalization of the model without losing accuracy. Different from centralized learning, each PV station trains its local model and only shares model parameters rather than original data to ensure privacy. Considering the framework robustness and training efficiency, the global model is aggregated distributedly to avoid central node failure. By designing the asynchronous update strategy, the communication overhead and training time are greatly reduced. Both experimental and numerical simulation results show that the global model obtained by the proposed method successfully diagnoses the types of faults contained in the data of all participants, and reaches the accuracy of centralized training. The performance of the proposed method in convergence speed, communication overhead, and training time is better than centralized federated learning.

APPENDIX A PROOF OF THEOREM 1

According to (15)-(17), we have:

$$G_t([e]) - G_t([e-1]) = -l_r \sum_{i \in L} \frac{d_i}{D} \nabla F_i(w_t^i([e-1])) \quad (22)$$

Considering the (18) and (22), we have:

$$\begin{aligned} & \|G_t([e]) - z_t([e])\| \\ &= \|G_t([e-1]) - l_r \sum_{i \in L} \frac{d_i}{D} \nabla F_i(w_t^i([e-1])) \\ &\quad - z_t([e-1]) + l_r \nabla F(z_t([e-1]), t)\| \\ &= \|G_t([e-1]) - l_r \sum_{i \in L} \frac{d_i}{D} \nabla F_i(w_t^i([e-1])) \\ &\quad - z_t([e-1]) + l_r \sum_{i \in L} \frac{d_i}{D} \nabla F_i(z_t([e-1]), t)\| \\ &= \|G_t([e-1]) - z_t([e-1]) \\ &\quad - l_r \sum_{i \in L} \frac{d_i}{D} (\nabla F_k(w_t^i([e-1])) - \nabla F_k(z_t([e-1])))\| \\ &\leq \|G_t([e-1]) - z_t([e-1])\| \\ &\quad + l_r \sum_{i \in L} \frac{d_i}{D} \|(\nabla F_k(w_t^i([e-1])) - \nabla F_k(z_t([e-1])))\| \end{aligned} \quad (23a)$$

$$\begin{aligned} &\leq \|G_t([e-1]) - z_t([e-1])\| \\ &\quad + l_r H \sum_{i \in L} \frac{d_i}{D} \|w_t^i([e-1]) - z_t([e-1])\| \end{aligned} \quad (23b)$$

$$\begin{aligned} &\leq \|G_t([e-1]) - z_t([e-1])\| \\ &\quad + l_r H \sum_{i \in L} \frac{d_i}{D} \frac{\delta_k}{H} ((l_r H + 1)^{e-1} - 1) \end{aligned} \quad (23c)$$

$$\begin{aligned} &= \|G_t([e-1]) - z_t([e-1])\| \\ &\quad + l_r \sum_{i \in L} \frac{d_i}{D} \delta_k ((l_r H + 1)^{e-1} - 1) \end{aligned}$$

where (23a) is from triangle inequality, (23b) is obtained because $F_k(\cdot)$ is H -smooth from Assumption 1, and (23c) is based on Lemma 2 in [36].

Since $\sum_{i \in L} \frac{d_i}{D} \leq \left(\sum_{i \in L} \frac{d_i}{D} + \sum_{j \notin L} \frac{d_j}{D} \right) = 1$ and $\delta_k \leq \bar{\delta}$, we have:

$$\begin{aligned} & \|G_t([e]) - z_t([e])\| \\ &\leq \|G_t([e-1]) - z_t([e-1])\| \\ &\quad + l_r \sum_{i \in L} \frac{d_i}{D} \delta_k ((l_r H + 1)^{e-1} - 1) \\ &\leq \|G_t([e-1]) - z_t([e-1])\| + l_r \bar{\delta} ((l_r H + 1)^{e-1} - 1) \end{aligned} \quad (24)$$

Thus, we have:

$$\begin{aligned} & \|G_t([e]) - z_t([e])\| - \|G_t([e-1]) - z_t([e-1])\| \\ &\leq l_r \bar{\delta} ((l_r H + 1)^{e-1} - 1) \end{aligned} \quad (25)$$

According to the definition of the auxiliary model $z_t([e])$, we have $\|z_t([0]) - G_t([0])\| = 0$. By summing up (25) over $e \in (0, \tau]$, we have:

$$\begin{aligned} & \|G_t([e]) - z_t([e])\| \\ &= \sum_{i=1}^e (\|G_t([i]) - z_t([i])\| - \|G_t([i-1]) - z_t([i-1])\|) \\ &\leq l_r \bar{\delta} \sum_{i=1}^e ((l_r H + 1)^{i-1} - 1) \\ &= l_r \bar{\delta} \frac{(1 - (l_r H + 1)^e)}{-l_r H} - l_r \bar{\delta} z \\ &= \frac{\bar{\delta}}{H} ((l_r H + 1)^e - 1) - l_r \bar{\delta} z \\ &= \bar{u}(e) \end{aligned} \quad (26)$$

Considering that $F(w, t)$ in Assumption 1 is ρ -Lipschitz, we have:

$$\begin{aligned} F(G_t([e])) - F(z_t([e])) &\leq \|F(G_t([e])) - F(z_t([e]))\| \\ &\leq \rho \|G_t([e]) - z_t([e])\| \\ &\leq \rho \bar{u}(e) \end{aligned} \quad (27)$$

Thus, Theorem 1 is proved.

APPENDIX B PROOF OF THEOREM 2

First, we analyze the convergence of $\|z_t(e+1) - G^*\|^2$ when $l_r \leq \frac{1}{H}$:

$$\begin{aligned} & \|z_t(e+1) - G^*\|^2 \\ &= \|z_t([e]) - l_r \nabla F(z_t([e])) - G^*\|^2 \\ &= \|z_t([e]) - G^*\|^2 - 2l_r \nabla F(z_t([e]))^T (z_t([e]) - G^*) \\ &\quad + l_r^2 \|\nabla F(z_t([e]))\|^2 \end{aligned} \quad (28a)$$

$$\begin{aligned} &< \|z_t([e]) - G^*\|^2 - l_r \frac{\|\nabla F(z_t([e]))\|^2}{H} + l_r^2 \|\nabla F(z_t([e]))\|^2 \end{aligned} \quad (28b)$$

$$\begin{aligned} &= \|z_t([e]) - G^*\|^2 - l_r \left(\frac{1}{H} - l_r \right) \|\nabla F(z_t([e]))\|^2 \\ &\leq \|z_t([e]) - G^*\|^2 \end{aligned} \quad (28c)$$

where (28a) is from expanding the squared norm, (28b) is based on lemma 3.5 in [38].

Then, considering that $F(\cdot)$ is H -smooth, we derive the upper bound of $F(z_t([e+1])) - F(z_t([e]))$:

$$\begin{aligned} & F(z_t([e+1])) - F(z_t([e])) \\ &\leq \nabla F(z_t([e]))^T (z_t([e+1]) - z_t([e])) \\ &\quad + \frac{H}{2} \|z_t([e+1]) - z_t([e])\|^2 \end{aligned} \quad (29a)$$

$$\begin{aligned} &\leq -l_r \nabla F(z_t([e]))^T \nabla F(z_t([e])) + \frac{H l_r^2}{2} \|\nabla F(z_t([e]))\|^2 \end{aligned} \quad (29b)$$

$$\begin{aligned} &\leq -l_r \left(1 - \frac{H l_r}{2} \right) \|\nabla F(z_t([e]))\|^2 \end{aligned} \quad (29c)$$

where, (29a) is based on lemma 3.4 in [38].

We define $\theta_t([e]) = F(z_t([e])) - F(G^*)$. Obviously, we have $\theta_t([e]) > 0$ for any round t and epoch e . From (29), we obtain:

$$\theta_t([e+1]) \leq \theta_t([e]) - l_r \left(1 - \frac{Hl_r}{2}\right) \|\nabla F(z_t([e]))\|^2 \quad (30)$$

Based on the convexity condition and Cauchy-Schwarz inequality, we have:

$$\begin{aligned} \theta_t([e]) &= F(z_t([e])) - F(G^*) \leq \nabla F(z_t([e]))^T (z_t([e]) - G^*) \\ &\leq \|\nabla F(z_t([e]))\| \|z_t([e]) - G^*\| \end{aligned} \quad (31)$$

Recalling (23) and $0 < \theta_t([e+1]) < \theta_t([e])$, we define $\chi \triangleq \min_t \frac{1}{\|z_t([0]) - G^*\|^2}$. Combining (30) and (31), and dividing both sides by $\theta_t([e+1])\theta_t([e])$, we have:

$$\frac{1}{\theta_t([e+1])} - \frac{1}{\theta_t([e])} \geq \frac{\chi l_r \left(1 - \frac{Hl_r}{2}\right) \theta_t([e])}{\theta_t([e+1])} \geq \chi l_r \left(1 - \frac{Hl_r}{2}\right) \quad (32)$$

$$\begin{aligned} \frac{1}{\theta_t([\tau])} - \frac{1}{\theta_t([0])} &= \sum_{e=0}^{\tau-1} \left(\frac{1}{\theta_t([e+1])} - \frac{1}{\theta_t([e])} \right) \\ &\geq \tau \chi l_r \left(1 - \frac{Hl_r}{2}\right) \end{aligned} \quad (33)$$

By summing up all the rounds, for $i = 1, 2, \dots, t$ and defining $T = t\tau$, we have:

$$\begin{aligned} &\sum_{i=1}^t \left(\frac{1}{\theta_i([\tau])} - \frac{1}{\theta_i([0])} \right) \\ &= \frac{1}{\theta_t([\tau])} - \frac{1}{\theta_1([0])} - \sum_{i=1}^{t-1} \left(\frac{1}{\theta_{i+1}([0])} - \frac{1}{\theta_i([\tau])} \right) \end{aligned} \quad (34a)$$

$$\geq T \chi l_r \left(1 - \frac{Hl_r}{2}\right) \quad (34b)$$

For the summation term in (34a), we have:

$$\begin{aligned} &\sum_{i=1}^{t-1} \left(\frac{1}{\theta_{i+1}([0])} - \frac{1}{\theta_i([\tau])} \right) \\ &= \sum_{i=1}^{t-1} \frac{\theta_i([\tau]) - \theta_{i+1}([0])}{\theta_i([\tau])\theta_{i+1}([0])} \\ &= \sum_{i=1}^{t-1} \frac{F(z_i([\tau])) - F(z_{i+1}([0]))}{\theta_i([\tau])\theta_{i+1}([0])} \\ &= \sum_{i=1}^{t-1} \frac{F(z_i([\tau])) - F(G_i([\tau]))}{\theta_i([\tau])\theta_{i+1}([0])} \end{aligned} \quad (35a)$$

$$\geq \sum_{i=1}^{t-1} \frac{-\rho\bar{u}(\tau)}{\theta_i([\tau])\theta_{i+1}([0])} \quad (35b)$$

$$\geq -(t-1) \frac{-\rho\bar{u}(\tau)}{\varepsilon^2} \quad (35c)$$

where (35a) is based on the definition of $z_t([e])$, and (35b) is derived by Theorem 1.

Combining (34b), (34c) and (35c), we have:

$$\frac{1}{\theta_t([\tau])} - \frac{1}{\theta_1([0])} = T \chi l_r \left(1 - \frac{Hl_r}{2}\right) - (t-1) \frac{-\rho\bar{u}(\tau)}{\varepsilon^2} \quad (36)$$

Based on Condition 4 in Theorem 2, we have $F(G_t) - F(G^*) \geq \varepsilon$. By combining Theorem 1 and $\theta_t(\tau) \geq \varepsilon$, we can derive:

$$\begin{aligned} &\frac{1}{F(G_t) - F(G^*)} - \frac{1}{\theta_t([\tau])} \\ &= \frac{F(z_t([\tau])) - F(G_t)}{(F(G_t) - F(G^*))\theta_t([\tau])} \\ &\geq \frac{-\rho\bar{u}(\tau)}{(F(G_t) - F(G^*))\theta_t([\tau])} \\ &\geq -\frac{\rho\bar{u}(\tau)}{\varepsilon^2} \end{aligned} \quad (37)$$

Summing up (36) and (37), we have:

$$\begin{aligned} \frac{1}{F(G_t) - F(G^*)} - \frac{1}{\theta_1([0])} &\geq T \chi l_r \left(1 - \frac{Hl_r}{2}\right) - t \frac{\rho\bar{u}(\tau)}{\varepsilon^2} \\ &= T \left(\chi l_r \left(1 - \frac{Hl_r}{2}\right) - \frac{\rho\bar{u}(\tau)}{\tau \varepsilon^2} \right) \\ &= T (\chi\varphi - \zeta) \end{aligned} \quad (38)$$

Based on Condition 2 in Theorem 2 and $\theta_1(0) > 0$, we have:

$$\begin{aligned} F(G_t) - F(G^*) &= \frac{1}{T(\chi\varphi - \zeta) + \frac{1}{\theta_1(0)}} \\ &\leq \frac{1}{T(\chi\varphi - \zeta)} \end{aligned} \quad (39)$$

Thus, Theorem 2 is proved.

ACKNOWLEDGMENTS

This work was supported by the National Key Research and Development Program of China (Grant No. 2018YFB1702300), and in part by the NSF of China (Grants No. 61731012, 62103265, 62122065 and 92167205).

REFERENCES

- [1] E. Papadis and G. Tsatsaronis, "Challenges in the decarbonization of the energy sector," *Energy*, vol. 205, p. 118025, 2020.
- [2] S. Saranchimeg and N. K. Nair, "A novel framework for integration analysis of large-scale photovoltaic plants into weak grids," *Applied Energy*, vol. 282, p. 116141, 2021.
- [3] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin, "Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and i-v characteristics," *Applied Energy*, vol. 204, pp. 912–931, 2017.
- [4] M. López-Escalante, E. Navarrete-Astorga, M. Gabás Perez, J. Ramos-Barrado, and F. Martín, "Photovoltaic modules designed for architectural integration without negative performance consequences," *Applied Energy*, vol. 279, p. 115741, 2020.
- [5] T. Takashima, J. Yamaguchi, K. Otani, T. Oozeki, K. Kato, and M. Ishida, "Experimental studies of fault location in pv module strings," *Solar Energy Materials and Solar Cells*, vol. 93, pp. 1079–1082, 2009.
- [6] M. U. Saleh, C. Deline, E. J. Benoit, S. R. Kingston, J. B. Harley, C. M. Furse, and M. A. Scarpulla, "Detection and localization of damaged photovoltaic cells and modules using spread spectrum time domain reflectometry," *IEEE Journal of Photovoltaics*, vol. 11, no. 1, pp. 195–201, 2021.
- [7] L. Chen and X. Wang, "Adaptive fault localization in photovoltaic systems," *IEEE Transactions on Smart Grid*, vol. 9, no. 6, pp. 6752–6763, 2018.
- [8] B. P. Kumar, D. S. Pillai, N. Rajasekar, M. Chakkarapani, and G. S. Ilango, "Identification and localization of array faults with optimized placement of voltage sensors in a pv system," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 7, pp. 5921–5931, 2021.

- [9] S. Saranchimeg and N. K. Nair, "A novel framework for integration analysis of large-scale photovoltaic plants into weak grids," *Applied Energy*, vol. 282, p. 116141, 2021.
- [10] J. Feng, Y. Yao, S. Lu, and Y. Liu, "Domain knowledge-based deep-broad learning framework for fault diagnosis," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 4, pp. 3454–3464, 2021.
- [11] Z. Yi and A. H. Etemadi, "Line-to-line fault detection for photovoltaic arrays based on multiresolution signal decomposition and two-stage support vector machine," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 11, pp. 8546–8556, 2017.
- [12] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, "Federated learning for data privacy preservation in vehicular cyber-physical systems," *IEEE Network*, vol. 34, no. 3, pp. 50–56, 2020.
- [13] S. Savazzi, M. Nicoli, and V. Rampa, "Federated learning with cooperating devices: A consensus approach for massive iot networks," *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4641–4654, 2020.
- [14] Gaofeng, L. Zhu, J. Wu, C. Shen, L. Zhou, and Q. Lin, "Blockchain-based federated learning for intelligent control in heavy haul railway," *IEEE Access*, vol. 8, pp. 176 830–176 839, 2020.
- [15] S. Mao, J. Wu, L. Liu, D. Lan, and A. Taherkordi, "Energy-efficient cooperative communication and computation for wireless powered mobile-edge computing," *IEEE Systems Journal*, pp. 1–12, 2020.
- [16] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in pv arrays: Types, detection, and mitigation techniques," *IEEE Journal of Photovoltaics*, vol. 5, no. 3, pp. 982–997, 2015.
- [17] M. Kumar and A. Kumar, "Experimental validation of performance and degradation study of canal-top photovoltaic system," *Applied Energy*, vol. 243, pp. 102–118, 2019.
- [18] N. Belhaouas, M.-S. A. Cheikh, P. Agathoklis, M.-R. Oularbi, B. Amrouche, K. Sedraoui, and N. Djilali, "Pv array power output maximization under partial shading using new shifted pv array arrangements," *Applied Energy*, vol. 187, pp. 326–337, 2017.
- [19] N. Sapountzoglou, J. Lago, B. De Schutter, and B. Raison, "A generalizable and sensor-independent deep learning method for fault detection and location in low-voltage distribution grids," *Applied Energy*, vol. 276, p. 115299, 2020.
- [20] M. Heinrich, S. Meunier, A. Samé, L. Quéval, A. Darga, L. Oukhelou, and B. Multon, "Detection of cleaning interventions on photovoltaic modules with machine learning," *Applied Energy*, vol. 263, p. 114642, 2020.
- [21] J.-M. Huang, R.-J. Wai, and G.-J. Yang, "Design of hybrid artificial bee colony algorithm and semi-supervised extreme learning machine for pv fault diagnoses by considering dust impact," *IEEE Transactions on Power Electronics*, vol. 35, no. 7, pp. 7086–7099, 2020.
- [22] Z. Chen, H. Yu, L. Luo, L. Wu, Q. Zheng, Z. Wu, S. Cheng, and P. Lin, "Rapid and accurate modeling of pv modules based on extreme learning machine and large datasets of i-v curves," *Applied Energy*, vol. 292, p. 116929, 2021.
- [23] Y. Zhao, R. Ball, J. Mosesian, J.-F. de Palma, and B. Lehman, "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays," *IEEE Transactions on Power Electronics*, vol. 30, no. 5, pp. 2848–2858, 2015.
- [24] S. Lu, T. Sirojan, B. T. Phung, D. Zhang, and E. Ambikairajah, "Dadegan: An effective methodology for dc series arc fault diagnosis in photovoltaic systems," *IEEE Access*, vol. 7, pp. 45 831–45 840, 2019.
- [25] N. Ding, Z. Fang, and J. Huang, "Optimal contract design for efficient federated learning with multi-dimensional private information," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 1, pp. 186–200, 2021.
- [26] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," *arXiv:1602.05629*.
- [27] Y. Liu, J. J. Q. Yu, J. Kang, D. Niyato, and S. Zhang, "Privacy-preserving traffic flow prediction: A federated learning approach," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7751–7763, 2020.
- [28] Y. Gao, L. Liu, B. Hu, T. Lei, and H. Ma, "Federated region-learning for environment sensing in edge computing system," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 4, pp. 2192–2204, 2020.
- [29] S. Lee and D.-H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," *IEEE Transactions on Industrial Informatics*, pp. 1–1, 2020.
- [30] W. Zhang, T. Zhou, Q. Lu, X. Wang, C. Zhu, H. Sun, Z. Wang, S. K. Lo, and F.-Y. Wang, "Dynamic fusion-based federated learning for covid-19 detection," *IEEE Internet of Things Journal*, pp. 1–1, 2021.
- [31] W. Wu, L. He, W. Lin, R. Mao, C. Maple, and S. Jarvis, "Safa: A semi-asynchronous protocol for fast federated learning with low overhead," *IEEE Transactions on Computers*, vol. 70, no. 5, pp. 655–668, 2021.
- [32] R. Miceli, A. Orioli, and A. Di Gangi, "A procedure to calculate the i-v characteristics of thin-film photovoltaic modules using an explicit rational form," *Applied Energy*, vol. 155, pp. 613–628, 2015.
- [33] Z. Chen, Y. Chen, L. Wu, S. Cheng, and P. Lin, "Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions," *Energy Conversion and Management*, vol. 198, p. 111793, 2019.
- [34] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv:1412.6980*.
- [35] Y. Liu, S. Garg, J. Nie, Y. Zhang, Z. Xiong, J. Kang, and M. S. Hossain, "Deep anomaly detection for time-series data in industrial iot: A communication-efficient on-device federated learning approach," *IEEE Internet of Things Journal*, vol. 8, no. 8, pp. 6348–6358, 2021.
- [36] S. Wang, T. Tuor, T. Salonidis, K. K. Leung, C. Makaya, T. He, and K. Chan, "Adaptive federated learning in resource constrained edge computing systems," *IEEE Journal on Selected Areas in Communications*, vol. 37, no. 6, pp. 1205–1221, 2019.
- [37] W. Wu, L. He, W. Lin, and R. Mao, "Accelerating federated learning over reliability-agnostic clients in mobile edge computing systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 7, pp. 1539–1551, 2021.
- [38] S. Bubeck, "Convex optimization: Algorithms and complexity," *Foundations & Trends in Machine Learning*, vol. 8, pp. 231–357, 2014.
- [39] W. Xia, W. Wen, K.-K. Wong, T. Q. Quek, J. Zhang, and H. Zhu, "Federated-learning-based client scheduling for low-latency wireless communications," *IEEE Wireless Communications*, vol. 28, no. 2, pp. 32–38, 2021.
- [40] S. Chen, H. Wen, J. Wu, W. Lei, W. Hou, W. Liu, A. Xu, and Y. Jiang, "Internet of things based smart grids supported by intelligent edge computing," *IEEE Access*, vol. 7, pp. 74 089–74 102, 2019.
- [41] W. Lei, H. Wen, J. Wu, and W. Hou, "Maddpg-based security situational awareness for smart grid with intelligent edge," *Applied Sciences*, vol. 11, no. 7, 2021.