electronics

MDPI

Review

Federated Learning for Condition Monitoring of Industrial Processes: A Review on Fault Diagnosis Methods, Challenges, and Prospects

Tarek Berghout [1], Mohamed Benbouzid [2,3,*], Toufik Bentrcia [1], Wei Hong Lim [4] and Yassine Amirat [5]

[1] Laboratory of Automation and Manufacturing Engineering, University of Batna 2, Batna 05000, Algeria
[2] Institut de Recherche Dupuy de Lôme (UMR CNRS 6027), University of Brest, 29238 Brest, France
[3] Logistics Engineering College, Shanghai Maritime University, Shanghai 201306, China
[4] Faculty of Engineering, Technology and Built Environment, UCSI University, Kuala Lumpur 56000, Malaysia
[5] ISEN Yncréa Ouest, L@bISEN, 29200 Brest, France
* Correspondence: mohamed.benbouzid@univ-brest.fr

Abstract: Condition monitoring (CM) of industrial processes is essential for reducing downtime and increasing productivity through accurate Condition-Based Maintenance (CBM) scheduling. Indeed, advanced intelligent learning systems for Fault Diagnosis (FD) make it possible to effectively isolate and identify the origins of faults. Proven smart industrial infrastructure technology enables FD to be a fully decentralized distributed computing task. To this end, such distribution among different regions/institutions, often subject to so-called data islanding, is limited to privacy, security risks, and industry competition due to the limitation of legal regulations or conflicts of interest. Therefore, Federated Learning (FL) is considered an efficient process of separating data from multiple participants to collaboratively train an intelligent and reliable FD model. As no comprehensive study has been introduced on this subject to date, as far as we know, such a review-based study is urgently needed. Within this scope, our work is devoted to reviewing recent advances in FL applications for process diagnostics, while FD methods, challenges, and future prospects are given special attention.

Keywords: condition monitoring; fault detection; fault diagnosis; federated learning

Citation: Berghout, T.; Benbouzid, M.; Bentrcia, T.; Lim, W.H.; Amirat, Y. Federated Learning for Condition Monitoring of Industrial Processes: A Review on Fault Diagnosis Methods, Challenges, and Prospects. *Electronics* **2023**, *12*, 158. https://doi.org/10.3390/electronics12010158

Academic Editor: Davide Astolfi

Received: 7 December 2022
Revised: 22 December 2022
Accepted: 23 December 2022
Published: 29 December 2022

1. Introduction

With the tremendous increase in the degree of diversification and decentralization within different computation platforms, new strategies have emerged in recent years to cope with severe constraints imposed on the elaboration of reliable machine-learning models. Given this, federated learning can be seen as a compromise between central models and decentralized data, which is of paramount importance in many fields, such as industrial plants. This explains why FL-enabling technologies in smart infrastructures, i.e., software, hardware, algorithms, and platforms, can efficiently handle lack-of-trust and privacy issues when sharing sensitive data [1–3]. FL is a learning process designed to accommodate such circumstances, where data sharing, in general, is subject to conflicts of interest. Indeed, its main plans are to improve the generalization of local learning models while only sharing the learning parameters of the models themselves. This review study was initiated with the aim of offering pertinent guidelines and helping researchers to pursue a path of countermeasure that could provide the best results (immune and accurate condition monitoring FL system). Accordingly, to ensure that its outlines are clear, this Introduction presents the context and main motivations leading to the current review-based study. More specifically, it aims to emphasize the research-based methodology followed in this work in addition to basic surveys and reviews of FL. In addition, it describes the major contributions and outlines the work. Thus, the main objective of the Introduction is to gain an overview of FL evolution while determining the best framework of the associated concepts in the analysis of FD works in the following sections.

1.1. Motivation

Condition monitoring systems play a crucial role in ensuring the continuous operation of production processes by minimizing downtimes. Using detection, diagnostic, and prognostic systems, CBM tasks are precisely planned, and systems are repaired at the right time while maximizing productivity [4]. The geographical distribution of industrial regions/institutions has led to so-called smart infrastructures adopting cyberphysical connectivity through specific networking and protocols [5]. Smart infrastructures generally connect edge devices, i.e., participants/clients, that are typically data-intensive and subject to a conflict of interest policy. Recent advances in data-driven methods have made machine-learning systems very successful in CM. In this context, the need to improve the generalization of machine-learning models without data sharing has yielded FL [6]. Both FL and FD have experienced many challenges; for instance, FL is prone to performance deterioration due to some data-complexity characteristics related to its volume, velocity, and variety. It could also be related to the type of edge devices, communication and networking methods, and protocols, as will be explained in Section 2. Moreover, FL algorithms or the learning process itself may be subject to security and privacy breaches by cyberthreats such as intrusions, for instance [7,8]. Wherefore, FD learning systems are mainly constrained to data availability, complexity, and drift. Motivated by the importance of both FD and FL in smart infrastructures CM, we are interested in providing a rigorous overview of FL in diagnostic systems while focusing on recent advances, challenges, and future prospects.

1.2. Methodology

The year 2016 can be considered the launch time of the FL paradigm [6]. Accordingly, in order to cover the majority of pertinent contributions dealing with the exploitation of FL in FD, we started our research/paper-collection process in the same year until September 2022, when this review study began. It should be mentioned that our research was carried out on the Google Scholar search engine, where we targeted comprehensive reviews, surveys, and research papers published each year separately. For each year, as addressed in Figure 1, we collected all works related to the context of the studied theme. Roughly speaking, we restricted the investigation to only the first eight pages of search results as they contain the main works, and beyond this limit, no relevant papers are encountered. Furthermore, high-quality papers published in reputable journals indexed in different well-known databases were addressed. As a result, the search process was conducted using keywords related to both FL and FD.

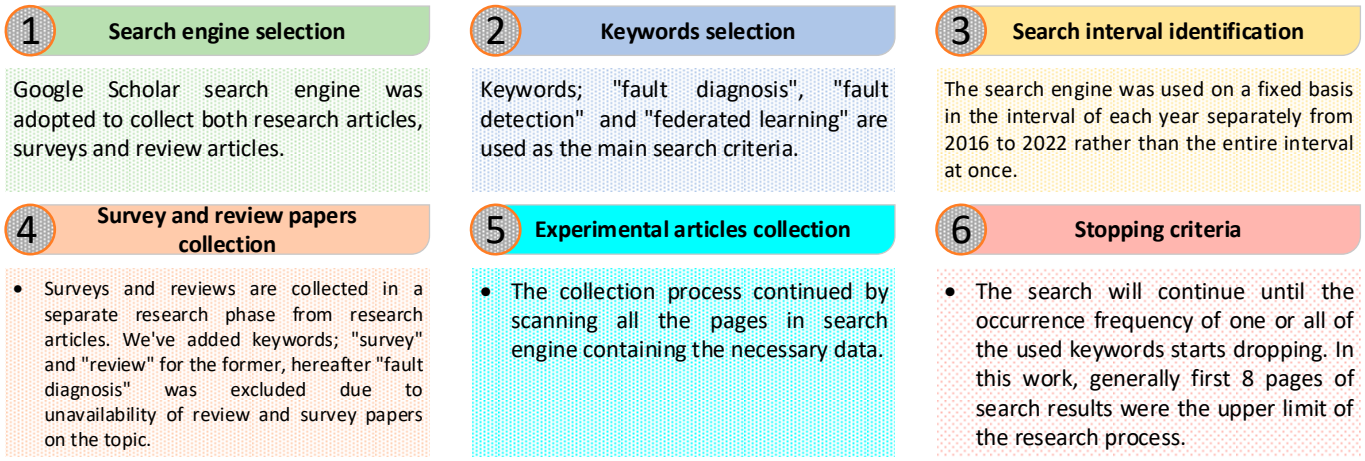

Figure 1. Methodology of paper collection.

First, this work was conducted to inspire important ideas about the field and introduce new input, and we scrutinized in-depth comprehensive studies with regard to the context of the theme of our work. Unfortunately, there is a significant lack of reviews and survey

papers on the aforementioned subject. Given this, instead, we collected all comprehensive studies, including reviews and surveys encompassing FL in a general manner only. Therefore, we kept only papers satisfying this criterion, which is related to terms such as Internet of Things (IoT), distributed environments, and edge and fog computing. We also excluded FL papers handling areas other than FL in general, e.g., healthcare, smart cities, vehicular IoT, natural language processing, and digital twins. Under these conditions, we estimated that the conclusions obtained from the general examination of FL would be more appropriate to some extent to the FD theme. Thus, the selected papers were sorted chronologically and analyzed. The analysis findings were used to provide insights for the synthesis study. During the collection process, it was noticed that these studies began about three years after the appearance of FL terminology. This means that the studied papers in this work ranged from 2019 to 2022, with a total of 18 examined papers.

The next step was to collect the research papers written on both FL and FD fields. We mainly followed a similar methodology using similar keywords from the year 2016. This time, related papers started appearing in 2020 and grew in number moderately during the next two years. We made sure that all papers, including targeted keywords, were collected, resulting in a total of around 19 research papers.

After that, and in order to acquire knowledge about advanced FL technology, the review papers were analyzed separately through information extraction using well-defined criteria, such as the review topic, the taxonomy used to classify FL methods, shortcomings of FL methods, and the types of discussed applications. Meanwhile, when analyzing research papers, criteria were considered that were tightly related to the main problem, the studied system/dataset, and the proposed solutions.

1.3. Related Surveys and Reviews: A General Context

It is noteworthy that the review papers were analyzed according to the publication year. Thus, classification, in this case, was also achieved based on the same criterion. This can be justified by observing the evolution of FL surveys in the context of algorithmic complexity, uncovered challenges, and addressed problems. In this work, we paid attention to remedied FL challenges so that we could carry out a meaningful analysis of the provided works in connection to FD. Accordingly, we scanned all FL reviews and created the most important list of challenges, namely, system heterogeneity (SyH), statistical heterogeneity (StH), privacy concerns (PrC), communication efficiency (CoE) (see Section 2), algorithmic architecture and automaticity, latency, massive distribution, and connectivity [9–11]. However, since latency and connectivity belong to the theme of CoE and mass distribution is involved with StH, these challenges were reduced to only four elements, namely, SyH, StH, CoE, and PrC, which will be explained in Section 2.

1.3.1. Research in 2019

The authors of [12] released a review paper focusing on the main concepts of FL, such as key definitions and the general mathematical context. The review paper paid more attention to FL types in the context of data architectures, i.e., feature types and samples. Accordingly, FL was classified into two categories, including vertical and horizontal FL, based on participants' data partitioning. In view of this, works on FL in privacy preservation, distributed learning, edge computing, and federated database systems were analyzed. Additionally, the application of FL was also briefly discussed.

1.3.2. Research in 2020

The year 2020 has produced many more comprehensive studies than in 2019, where the privacy of FL models itself began to appear as responsible for security issues. For instance, in addition to the aforementioned FL categories and concepts established in previous work [12] and some FL characteristics related to previously mentioned FL challenges, work in [13] also looked at the description of available open-source FL frameworks, including the TensorFlow FL Framework (TFF) [14] and Federated AI Technology Enabler (FATE) made

available by the Webank team [15]. From this perspective, there is an interesting illustration of the evolution of FL in chronological order (see Figures 5 and 6 from [13]). Moreover, this work paid more attention to the challenges faced by the application of FL algorithms as well as their evolution. Additionally, it discussed applications of FL in IoT devices, industry, and healthcare. The review presented in [9] aimed to discuss the unique characteristics and challenges of FL by providing a broad picture of approaches and identifying several directions for future work. The review introduced in [16] studied FL in terms of security and PrC, where security and privacy issues' identification, evaluation, and documentation were tackled. In that respect, it listed security and privacy use cases and provided risk factors, helping researchers attain a clear view of possible research axes. It also presented an illustrative description of approaches, various implementations, and challenges and established a review of security and PrC. Additionally, it specified the most important security threats, including poisoning and bottlenecks, while inferencing that attacks require more effort.

1.3.3. Research in 2021

In 2021, there was profound development in understanding FL in greater detail, while PrC took a prominent place among other challenges. The work presented in [17] was specifically devoted to the analysis of privacy preservation in FL. Essentially, in addition to previous concepts and major challenges of FL, this review focused on presenting the latest advances in cryptographic, disturbance, anonymization, and privacy measures. These achievements were consolidated by proposing a general taxonomy to analyze privacy risks in FL in terms of various aspects, summarizing existing methods, and identifying future research directions. Similarly, the comprehensive study presented in [18] reviewed the development of FL and presented the current contributions according to a new classification of five aspects, namely, data partitioning, privacy, ML algorithms, communication architecture, and SyH. In addition, it detailed recent challenges and future opportunities for FL whilst considering the most important characteristics of FL methods and their application. The survey in [19] was dedicated to exposing advances in mitigating security and privacy attacks. To this extent, FL analysis targeted solutions, evaluations, and comparison methods to maintain privacy. Moreover, this survey analyzed drawbacks related to maintaining privacy on both server and client sides at the same time. Hence, the survey typology was designed to discuss privacy-enabled constraints and solutions while highlighting challenges and charting methods to achieve both types of protection. Another detailed review that investigates FL survey and research papers was showcased in [20]. This review presented FL research as a broad category of concepts, challenges, and applications. Moreover, it reviewed data partitioning, FL architectures, clustering techniques, and aggregation techniques' main aspects, i.e., CoE, SyH, SyH, and PrC, as well as different areas of application, in addition to discussing open issues and challenges in FL. In [21], the latest findings of FL implementation, recent developments, and insights into IoT networks were discussed. Additionally, special attention was given to FL opportunities in data sharing, data loading and caching, attack detection, localization, mobile audience sensing, and IoT privacy and security. Likewise, general aspects of FL in smart IoT infrastructures, such as healthcare, transportation, drones, smart cities, and industry, were also addressed in the context of networking aspects and protocols while summarizing tables of FL aspects, contributions, and limitations of the research papers were also provided. A prominent contribution of this review consists of providing the most important lessons learned on the topic, fundamental research challenges, and prospects of future research. The work in [10] provided a comprehensive picture of FL and its taxonomy, going through the most important classification of FL, challenges, vulnerability assessment approaches implications, limitations, and future scope. In the framework of FL classification, it delved into several types of classifications based on different aspects, including data partitioning, modeling methods, privacy level, communication architecture, and data availability. Additionally, it added two new challenges to the previous four challenges, i.e., CoE, SyH, SyH,

and PrC—namely, algorithmic architecture and automaticity. In [22], papers devoted to discussing FL in the context of user-generated data privacy in the IoT were underlined. Challenges related to the CoE, StH, and additional privacy issues associated with FL were also highlighted. The strengths and weaknesses of the different methods applied to FL were listed in this review paper, while future directions for preserving privacy on IoT applications were also taken into consideration. The review paper in [23] targeted FL challenges from the perspective of non-IID on parametric and non-parametric machine learning for both horizontal and vertical FL. It therefore enumerated the challenges, advantages, disadvantages, and future research directions of non-IID data in federated learning. The work in [24] focused on FL advances and introduced challenges in terms of maintaining privacy. It also featured the background, motivations, definitions, architectures, and applications of FL as a privacy-preservation paradigm. Finally, an extremely inclusive and deep review covering the finest multilevel classification was offered in [11]. This paper introduced a comprehensive study of FL, related concepts, technologies, and learning methods. It also depicted the applications and future orientations of communication and networking. On this basis, a three-level classification scheme was designed to categorize FL literature according to a high-level challenge (i.e., CoE, StH, SyH, PrC, client selection and scheduling, and service selection). After, each high-level challenge was categorized by itself into sub-level challenges, furnishing more details. Moreover, in each sub-level challenge, an accurate classification based on used techniques was provided, and guidelines for future research were given for each category.

1.3.4. Research in 2022

The year 2022 produced more detailed reviews of other FL challenges, such as StH, while PrC studies delved even further; for example, the work in [25] was concerned with recent progress in FL against evolving edge-computing standards. The review was focused on training with the employment of deep learning. Thus, the work was conducted to illustrate edge computing, deep learning, and key FL concepts, summarizing the contributions of the literature review and discussing FL architecture, challenges, and recent solutions. It also reported some examples of case studies related to drones and healthcare. Finally, it listed open issues and possible future research. The review paper in [26] was dedicated to analyzing and establishing the definition of non-IID data problems for FL while pointing out the challenges of this specific problem. Accordingly, the methods used to deal with this problem were classified with the aim of providing a comprehensive study to solve the problem of non-IID in FL. The results showed that non-IID data were the main reason for reduced FL performance and the harmful active participation of clients. The work in [27] attempted to review FL trends in edge computing by synthesizing and comparing FL algorithms, models, and frameworks while presenting different applications. In addition, datasets used for FL were specified, while current challenges were also discussed. In [28], various research on centralized FL, decentralized FL, and heterogeneous FL in the context of distributed environments were analyzed. Moreover, privacy-preservation analysis and data sharing, as well as confidentiality in the maintenance of the security framework, were elaborated. Finally, the review demonstrated the challenges, critical analysis, and parameters of the FL model.

Table 1 features a recapitulation of analyzed FL reviews and surveys. Generally speaking, all reviews followed a similar methodology when introducing their studies' papers. First, they addressed the main concepts and familiar topics related to FL, then delved into the classification of FL methods. Classification depended on the main goal of the review and some criteria. For example, when considering data partitioning, FL could be classified as vertical, horizontal, and transfer federated learning. However, addressing the challenges of FL is a necessary point that all reviews agreed on. In this scope, some reviews analyzed contributions dedicated to all challenges, and others delved into more detail on a specific topic, as already showcased in Table 1.

Table 1. Investigated review/survey papers of FL in a general context.

Year	Ref.	Most Important Contributions besides Generalities, e.g., Main Concepts, Classifications, Applications, and Future Prospects	Main Challenges Addressed
2019	[12]	• Classification of FL in two categories, including vertical and horizontal FL based on data features; • Analysis of FL works according to privacy preservation, distributed learning, edge computing, and federated database systems.	Challenges of FL in general
2020	[13]	• Discussion of FL challenges: system heterogeneity, StH, PrC, and communication efficiency; • Description of open-source TensorFlow and Federated AI Technology.	Challenges of FL in general
	[9]	• Discussion of FL challenges: system heterogeneity, StH, PrC, and communication efficiency.	Challenges of FL in general
	[16]	• Focused on security and PrC by identifying security and privacy issues, and evaluating and documenting them, respectively.	PrC
2021	[17]	• Analyzing privacy preservation in FL by presenting advanced cryptographic, disturbance, anonymization, and privacy measures.	PrC
	[18]	• Classification according to five aspects, namely, data partitioning, privacy, ML algorithms, communication architecture, and systems heterogeneity.	Challenges of FL in general
	[19]	• Discussion of mitigation security and privacy attacks.	PrC
	[20]	• Discussion of FL challenges: system heterogeneity, StH, PrC, and communication efficiency; • Discussion on data partitioning, FL architectures, and clustering techniques.	Challenges of FL in general
	[21]	• Discussion of FL opportunities in data sharing, data loading and caching, attack detection, localization, mobile audience sensing, and IoT privacy and security; • Discussion of general aspects in the context of networking aspects and protocols.	StH and PrC
	[10]	• Provision of different types of classifications based on different aspects, including data partitioning, modeling methods, privacy level, communication architecture, and data availability. • Addition of two more new challenges to the previous four challenges, namely, algorithmic architecture and automaticity.	Challenges of FL in general and PrC
	[22]	• Discussions of data privacy in the IoT; • Highlighting challenges related to the communications expensiveness, StH, and additional privacy issues associated with FL were also highlighted.	PrC
	[23]	• Targeted StH.	StH
	[24]	• Focused on advances in FL and addressed challenges in terms of maintaining privacy.	PrC
	[11]	• Focused on proving a deep taxonomy of FL.	Challenges of FL in general
2022	[25]	• Discussion of FL in the context of using deep-learning methods.	Challenges of FL in general
	[26]	• Analyzed FL works in context of non-IID.	StH
	[27]	• Discussion of FL in the context of edge computing.	Challenges of FL in general
	[28]	• Discussion of privacy preservation, data sharing, and confidentiality.	PrC

The chronological order shows that review papers started by studying FL methods from a general viewpoint. After, efforts were concentrated on each challenge separately, starting with the PrC of FL models themselves. Then, pure comprehensive studies appeared to deal with other challenges, such as StH. This means that StH is one of the primary concerns of FL models after their security design to protect from cyberattacks. In order

to bring attention to the most important studied challenges regarding FL, we created a pie chart, which is shown in Figure 2. This figure summarizes the important topics and main challenges of FL. It indicates that, before all, the privacy of the FL algorithm is necessary and should be carried out. Thenceforth, StH appeared to be very important in constructing a powerful FL model. One of the limitations we found in the literature is that there is a scarcity of comprehensive reviews, specifically in connection with other CoE and SyE challenges.

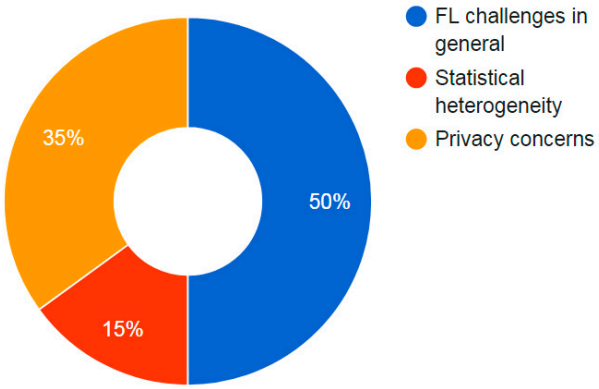

Figure 2. Main investigated types of FL challenges.

Based on these findings, FD tools developed based on FL algorithms will also be analyzed under these FL challenge criteria according to their importance. This indicates that such tools will be analyzed based on how they deal with the privacy protection of the FL model and a strong background of StH. They will also be analyzed on whether they address CoE and SyH.

1.4. Contributions

Apart from FL review and survey analysis in a general context while mainly focusing on the challenges in Section 1, this review paper is devoted to the following:

- Providing a brief background on the main concepts, classification, algorithms, and challenges of FL. This background is limited to the information needed for this review study.
- Analyzing FD methods in the context of the treated problems, FL algorithms, machine-learning algorithms, and datasets.
- On the basis of the obtained results, this review provides challenges facing the evolution of FL in the context of FD.
- This review also gives insight into the future prospects of FD applications in FD.

1.5. Review Outlines

This paper is organized as follows: in addition to the Introduction of Section 1, Section 2 is devoted to providing a general overview of FL. Accordingly, only necessary information in the context of our review-based study is revealed. Section 3 is allocated to reviewing recent progress in FD methods in the context of using FL for privacy preservation, listing the main challenges facing the evolution of FL in FD, and providing future directions for FD under FL. Section 4 concludes this work and gives a summary of future work to be carried out as a continuity of this study.

2. Overview of Federated Learning

This section is mainly reserved for the description of the necessary concepts and definitions related to our study. As this is already widely known in the literature, we have made it as short as possible and left only the important features that are necessary to understand the content of this work. Accordingly, Figure 3 will be used in this sec-

tion to discuss FL mechanisms in relation to the positions of FL challenges as well as its enabling technologies.

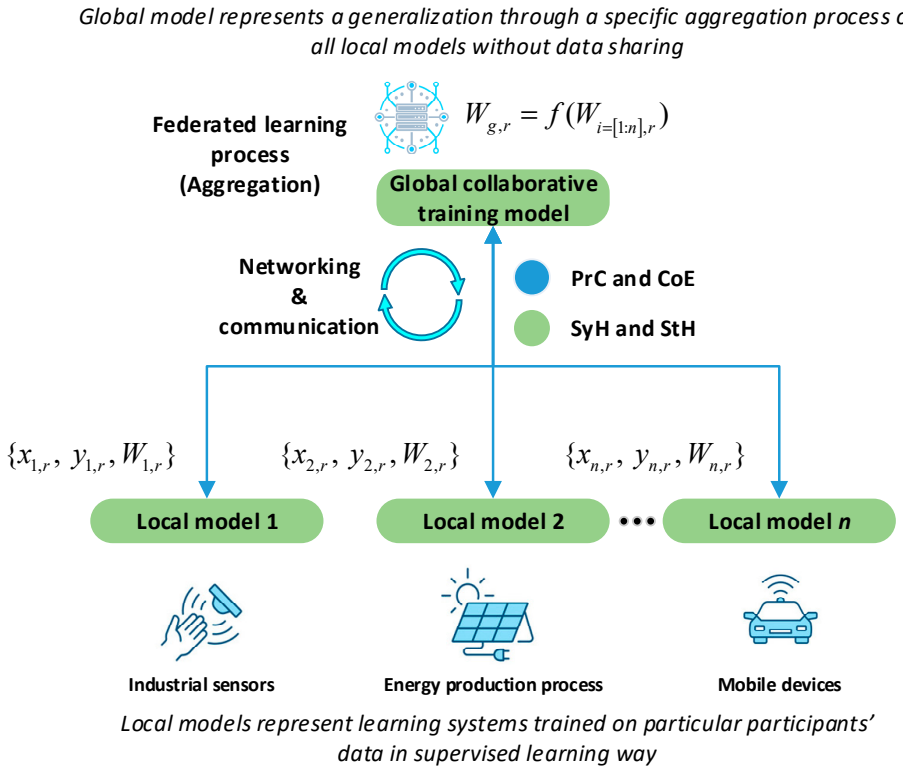

Figure 3. Overview of FL in a parallel-computing pool.

2.1. Overview

FL is a collaborative training method designed to improve the generalization of local machine-learning models by transferring their training parameters to a single global model while keeping data localized only in the local models. More precisely, FL is based on running an algorithm in a parallel-computing pool to perform aggregation of a set of different learning models while taking into account several aspects related to communication efficiency, system heterogeneity, StH, and PrC. FL is an online learning process that manipulates models iteratively and adaptively trains for a specific period of time, known as a training round, on a specific set of mini-batches of inputs and targets $\{x_{c,r}, y_{c,r}\}$ of a specific type of data, where $\{c, r\}$ refers to the client and training round, respectively. This process continues as long as training data keep flooding sequentially. Assume that $W_{c,r}$ are learning parameters obtained from participants at a specific training round from specific training model. FL's main mission, in this case, is to aggregate $W_{c,r}$ at each round while training to maximize the accuracy of local learning models by minimizing the losses l accordingly. As represented by (1) and (2), f denotes the main function of aggregation; g refers to the global model; and the loss function is presented by θ, which is defined according to training process constraints. $\widetilde{y}_{c,r}$ refers to estimated targets from each local model.

$$W_{g,r} = f(W_{c,r}) \tag{1}$$

$$l_{c,r} = \theta(y_{c,r}, \widetilde{y}_{c,r}) \tag{2}$$

The aggregation function f differs from one model to another. In fact, its design mainly depends on the treated subject and the main challenges to be solved. The most basic

aggregation function is based on federated averaging (FedAvg) of learning parameters, as described by (3).

$$W_{g,r} = \frac{1}{c} \sum_{i=1}^{c} (W_{c,r}) \tag{3}$$

2.2. Classification of FL Algorithms

FL algorithms can be classified according to many criteria, such as network topology, data availability, data partitioning, aggregation and optimization techniques, and open-source frameworks (see [16], Figure 2), depending on the main objectives of the study. In our review-based study, we are interested in the analysis of FL methods according to the type of data partitioning (Section 3 demonstrates that other classifications have not been given sufficient attention). So, the FL algorithms, in this case, are categorized under three types, namely, horizontal FL, vertical FL, and transfer FL. Vertical FL training scenarios involve local datasets sharing the same sample feature identification with different spaces in one or more of the following data features: data variation, velocity, and space volume [10]. Horizontal FL represents the case where local datasets might share a few feature-identification samples but the same feature space. Federated TL refers to different local datasets, where the challenge is the need to improve generalization from each other due to the advantages of data behaviors.

2.3. Challenges

As discussed earlier, FL is subject to different challenges, mainly related to SyH, StH, PrC, and CoE. Since this is very important for our study in this case, as the main FD methods will be analyzed and criticized according to these criteria, this section aims at describing these challenges.

SyH: FL networks are natively heterogeneous due to differences in many characteristics, such as connectivity, hardware, powering methods, communication, storage capacity and method, and node-computing capabilities. Additionally, not all nodes may be active at the same time due to the fact that some devices work periodically or may also be under maintenance. Devices may stop functioning at some point during FL; therefore, fault tolerance is enabled by these system-level properties [27]. In this context, the FL algorithms and the training mechanism must take this factor into account because it plays a vital role in balancing the scales of the training models via weighting.

StH: Typically, participants collect non-IID across the network with significant variations, leading to unique statistical structures among devices and their associated distributions. This non-IID nature adds some complexity in terms of modeling, analysis, and evaluation. As a result, multitasking and meta-learning enable personalized modeling for specific devices, which is often a more natural approach to dealing with StH [9].

PrC: Privacy is a major concern in FL networks. FL already relies on a privacy-protection policy by sharing only trained local models to obtain updates without data sharing. However, multiple participations in multiple training rounds can divulge sensitive information from data. For instance, a long short-term memory (LSTM) network with the ability of sequential learning can reveal information about data under several attack tests due to the forgetting mechanism. FL's mission, in this case, is to enhance privacy with tools such as differential privacy. However, these methods often provide privacy at the cost of degrading model performance or system efficiency [9].

CoE: Communication is a critical issue in FL networks, especially when associated with privacy. The massive number of devices driving massive communications will certainly slow down, in terms of magnitude, local computing with limited resources, such as bandwidth, energy, and power. As a result, it is very important to design CoE methods to reduce communication rounds and their sizes, respectively.

2.4. Enabling Technologies

Many technologies have been applied and used to improve FL further. As stated earlier, these technologies range from algorithms to hardware and software platforms. Hence, this subsection is devoted to the introduction of the most important and well-known technologies so far.

2.4.1. Algorithms

FedAvge initiates the first attempts of FL. Nevertheless, when it comes to other challenges, it seems that, in this case, it is designed to fit with accuracy and performance only, and it does not consider, for instance, the privacy, communication, or heterogeneity of data and systems. So, the following versions of FL algorithms released were oriented to such issues either separately or totally. In this context, this section will only briefly cover important aspects of some well-known FL variants, including FedProx for StH, FedPAQ for CoE, and TurboAggregate for both CoE and the security of FL models. As an example, FedProx [29] is designed to address heterogeneity primarily related to non-IID between participants. The main idea is to add mathematical expressions to the aggregation functions as a kind of proximal term for each local training parameter of each local model separately. This will force the update processes to approximate the global model and reduce the influence of StH with each update. Overall, FedProx is an FL philosophy targeting loss function convergence established under non-IID. In FedProx, all devices are weighted equally, the same as in FedAvg at the aggregation phase. Accordingly, the differences between participants (e.g., hardware) are not taken into account, i.e., SyH. FedPAQ [30] is a CoE aggregation process that offers periodic averaging instead of synchronizing updates with the server. FedPAQ involves a variety of local updates before any collaboration. The primary goal is to reduce communication overhead by allowing a subgroup of client devices to participate in training based on the device's connectivity to a wireless network free, idle, and accessible to a base station. Identically to FedAvg, the new global model of FedPAQ is computed as the average of the local models, which involves high complexity in strongly convex and non-convex parameters. TurboAggregate [31] is an aggregation algorithm implemented to handle both PrC and CoE aggregation. It is based on an aggregation of several groups of participants in a circular strategy. Initially, at the level of CoE, it iteratively divides the participants into several groups, sends them for global aggregation, and employs the updated model to adjust the other groups, respectively. It also includes a rest security feature by adding random variables as noise to the local models to preserve privacy. The noise is designed to be filtered when the update process is complete.

2.4.2. Software and Platforms

Fortunately, there are many FL open-source systems that support the training of machine-learning models, while the security of such an open platform ensures no leakage of either sensitive data or model parameters [24]. Indeed, a variety of open-source software platforms, such as Federated AI Technology Enabler (FATE), TensorFlow Federated (TFF), OpenMined, LEAF Benchmark, FedML, PaddleFL, and OpenFL, are widely used in different industrial contexts [24]. Institutions such as the AI department of WeBank, a digital bank in China; Carnegie Mellon University; and Google AI provide these open-source platforms to perform FL learning tasks, giving access to many privileges, such as decentralized learning and FL algorithms with different architectures (including deep learning, different FL datasets, etc.). These FL platforms also provide secure-computing protocols and privacy-preserving tools such as homomorphic encryption and multi-party computing.

2.4.3. Hardware

The algorithmic complexity of deep learning motivates accelerating hardware development to support FL computational and storage requirements. Indeed, the multiple layers of non-linear abstractions in deep learning lead to greater computational complexity. A massive number of training parameters and iterative learning operations and millions

of samples require more computational power, especially under FL. In this context, FL imposes the use of three categories of computational resources. These include Central Processing Units (CPUs), Graphics Processing Units (GPUs), and Application-Specific Integrated Circuits (ASICs) [25]. GPUs are effectively used for cloud computing due to the challenges of big data and the use of deep learning. In the case of FL, it is recommended to use GPUs when training global models that require tuning massive parameters, including potential data processing. For the resource allocations of typically lightweight peripheral devices that require fewer computing resources due to limited computing power, CPUs are very efficient. Meanwhile, ASICs are recommended to simplify the algorithmic complexity of the deep-learning model and speed up its application [25].

3. Works Related to Federated Learning for Fault Diagnosis

This section will deal with the review of papers falling exactly within the subject of the paper, which is FL for FD. It also presents the main challenges and future prospects.

3.1. Analysis

3.1.1. Research in 2020

In 2020, an FL algorithm was proposed to realize a three-level fault detection system for power terminals [32]. The main idea was to tackle expansive communication constraints by reducing message transmissions between the cloud and edge servers while protecting data privacy. It should be mentioned that the data used here were picked from a simulated power grid system with a cloud, 10 edge servers, and numerous terminals. It represents a linguistic description of the system health status recorded from a power terminal system log, so the FL problem is a language-processing topic. In this case, long short-term memory (LSTM), which is a better choice for sequential learning, was adopted to train the collaborative model on edge devices. Once LSTM was trained, the training parameters were sent to the local server for aggregation. It was stated that the model might be prone to overfitting due to data imbalance when receiving a lower number of fault samples. Therefore, the FL algorithm was designed to take these issues into account when using the FedAvg philosophy. Additionally, the log records of electrical terminals were expected to be massive, which would lead to higher transmission costs. Therefore, these records were further compressed with specific three-stage log compression, i.e., variable replacement, similarity calculation, and redundancy filtering, respectively, to ensure low latency.

3.1.2. Research in 2021

For the year 2021, an FL method for machine FD with dynamic validation and self-supervised learning was proposed while treating both data privacy and scarcity issues for different clients in [33]. In this context, a federated deep-learning network was proposed for global aggregation. Moreover, a dynamic-validation scheme was adopted to solve the problems of the StH of data. Two rotating machinery datasets, namely, the Case Western Reserve University (CWRU) [34] and Bogie bearing datasets, were involved. CWRU describes the vibration acceleration signals collected from the drive end of the motor, while the Bogie dataset was similarly collected from a multi-unit train bogie high-speed running system. A self-supervised data translation (augmentation) method was proposed to deal with problems caused by unbalanced or limited data patterns. FedAvg algorithms were adopted to train a collaborative model based on CNN architecture for further extraction and classification (i.e., supervised learning). A new aggregation methodology was proposed in [35] to provide greater privacy to FL algorithms in the fault diagnosis of rotating machinery. The main idea was to adopt a so-called quasi-cloud/edge/client-learning mechanism by involving sequential Kalman filter learning rules as an asynchronous fusion algorithm of learning parameters provided by edge clients. The effectiveness of the proposed scheme was verified by the data of rotating machinery, and the authors succeeded in using their own dataset retrieved from a specific rotating mechanical device. Pedestal loosening, broken fan blades, and rotor imbalance were the main considered failure modes

of the system. After a well-defined feature-extraction process of vibration signals obtained from eight sensors installed at different locations, a single hidden layer neural network with adjustable output weights only was involved in training edge models. Continuing the work presented in [35], the authors in [36] added new features to the learning scheme in addition to asynchronous Kalman filtering. A real-time customer identification method with the newly labeled samples obtained at unequal intervals was approved to enable clients to reach better performance. Additionally, the CWRU dataset [34] was also studied in addition to a previously studied one to ensure that their study could be generalized better. In [37], a fault-diagnosis method based on the FedAvg algorithm to collaboratively train a CNN model was proposed. Two main datasets, namely, CWRU [34] and Machinery Fault Database (MFD) [38], were involved in an attempt to provide an effective investigation of the FL model under limited data sharing and poor generalization of the model for mechanical equipment. This type of challenge in the context of FL falls under the theme of StH. In [39], an asynchronous FL method was proposed to solve problems related to resource constraints that fall within the field of CoE and system heterogeneity. This method enables efficient node selection for asynchronous updates based on local data distribution. In view of this, the main process led to reducing computational costs and improving communication and the efficiency of the FL model. The FL algorithm in this work adopted a multi-branch neural network (MBNN) as an optimization method for asynchronous FL. This method depended on penalization to select only the branches of interest (nodes important for training). In order to evaluate the proposed MBNN, two real datasets were involved, namely, the CWRU dataset [34] and the gearbox failure diagnostic dataset provided by the University of Connecticut [40,41]. In [42], and under computational cost and CoE criteria, an FL method that contains stacked sparse autoencoders (SSAE) and Siamese networks was proposed for the diagnosis of inter-turn short-circuit (ITSC) faults in permanent magnet synchronous motors (PMSM). SSAE was used for unsupervised feature extraction with an incomplete list of patterns, while Siamese networks were used for similarity analysis. An adaptive FL model was proposed in [43] to adjust the FL aggregation interval based on some feedback edge clients to target communication cost while ensuring model accuracy. Accordingly, a CNN with a dropout layer was trained by momentum gradient descent (MGD) to speed up the convergence rate and avoid overfitting. The efficiency of the proposed FL method was satisfactory when applied on a non-IID bearing fault (CWRU) dataset. A federated transfer-learning approach was investigated for machine fault diagnosis in [44]. A deep adversarial semi-supervised network was used for transfer learning, while it could be initialized differently depending on each particular client. The model was built with the intention of targeting SyH as the primary objective rather than StH. A deep network is a type of CNN used for supervised learning. The same CWRU [34] and Bogie datasets discussed earlier were used for model validation.

3.1.3. Research in 2022

In 2022, a federated transfer-learning method using the problem of StH and the unavailability of data caused by different working conditions and accelerated tests was proposed for the fault diagnosis of machines in [45]. The preceding distributions for bridging the domain gap that follows demonstrated information sharing across domains while maintaining data privacy. The training process was carried out involving a specific CNN architecture. Two datasets, namely CWRU [34] and CRACK, were implemented for validation. In [46], the authors proposed a federated deep-learning framework for fault detection in wind turbines. Accordingly, a multiscale residual attention network model was used for feature extraction from raw data. Two particular tasks resulted in this case. First, a multiscale residual learning process was exploited for spatial feature extraction from different sensors at different scales. Second, a feature-attention mechanism was dedicated to selecting only descriptive and important features strongly correlated to fault events, leading to an improved fault-detection process of a final universal approximator (i.e., a deep network designed for classification). From the provided algorithms, it appeared that

the standard FedAvg algorithm was used for the aggregation process to deal with the StH problem under large data unavailability. Two sets of condition-monitoring data from an actual wind turbine were used for the evaluation of the proposed model. The datasets were collected from a supervisory control and data-acquisition system installed to monitor two turbines. The first was a 3 MW direct drive wind turbine near the south coast of Ireland [47]. The second was acquired from a 1.5 MW direct-drive WT in northern China [48]. A deep-learning-based diagnostic method based on federated adversarial and transfer learning that enables domain adaptation was developed in [49]. Accordingly, a new federated minimax (FedMM) algorithm was also developed for global model aggregation that targets the problem of gradient drift caused by sample imbalance under StH. Complex deep-learning networks such as CNN were involved in training for the supervised approximation process. The model was verified under gas-insulated switchgear insulation faults with experiments consisting of laboratory and field clients. In [50], a new adaptive federated learning approach for Internet of Ships fault diagnosis was proposed. The objective of the method was to target PrC using the Paillier cryptographic communication scheme. Additionally, to deal with the StH of the harsh marine environment, an adaptive control algorithm was proposed to deal with the model-aggregation interval during the learning process while reducing cryptography computations and communications expensiveness. The CNN algorithm was adopted for the classification process while a CWRU was involved [34]. In [51], a stacking federated learning model was proposed for diagnosing short-circuit faults between turns in permanent magnet synchronous motors. Under connectivity criteria that fit into communication efficiency, a verification strategy was implemented for selecting clients. A dataset was created in this context to validate the client weightings. The dataset contained samples for normal operating conditions and fault patterns. Particle swarm optimization was involved in finding weights for clients engaged in the aggregation process. Compared to the standard FedAvg, the proposed model showed better performance in terms of data imbalance and communication costs, and local oscillations were avoided in the model. In [52], authors proposed a federated-learning-based diagnosis system for fault detection and identification in PV systems. To address the challenges of system and StH different computing capabilities and the amount of data in the PV station, an asynchronous decentralized federated learning (ADFL) mechanism was designed. The asynchronous update scheme helped in overcoming communication overhead while reducing training time. A specific weight mixing method proposed by the authors was used to perform model aggregation (see Equation (11) of [52]), while a CNN was the main training architecture. Experiments on a real PV generator consisting of two PV strings connected in parallel with 22 PV modules in series were performed to verify the efficiency of the proposed method, while short circuit faults, partial shading faults, and degradation faults were simulated, respectively. In [53], the authors proposed a fast FedAvg (FA-FedAvg) algorithm for fault diagnosis based on traditional FedAvg algorithms to improve model learning quality and speed. As a result, a client-selection weighting strategy was incorporated to conduct high-quality communication. Under latency criteria, an aggregation strategy based on the precision difference was proposed to reduce the number of iterations and accelerate the convergence of the learning model. Finally, the proposed algorithm was applied to the CWRU dataset [34] using a CNN architecture and compared with FedAvg and FedProx. In the work proposed in [54], more attention was paid to class imbalance when building their privacy-preserving federated learning framework for the fault diagnosis of decentralized wind turbines. In this framework and taking into account PrC, the authors built a biometric authorization mechanism to ensure that only legitimate access granted private data and served against cyberthreats. The federated learning process itself was built on a dual privacy-enhancement mechanism; namely, the gradient noise mechanism and the proportional parameter update strategy were used to enable privacy and security, while ResNet18 [55] was the main training model. Meanwhile, the gradient-based self-monitoring scheme was integrated for overall imbalance fault diagnosis. Vibration data that were collected from five biometric authentication wind farms located in Kangbao, Hebei

Province, China, were used to verify the effectiveness of the proposed framework. In [56], under the problem of StH and data unavailability, a CNN average shared layers through federated transfer learning (FTL-ASL) method for bearing fault diagnosis was proposed. Then, a modified FedAvg algorithm was adopted for aggregating feature layers from different diagnostic models, while custom layers were updated locally. A set of bearing datasets, i.e., Paderborn University (PUD) [57], Machinery Failure Prevention Technology (MFPT) [58], and CWRU [34], were used to validate the implemented approach. In [59], a federated transfer learning framework with deviation-based weighted federated average (D-WFA) was proposed for bearing fault diagnosis. As a result, a dynamic weighted-average algorithm based on maximum mean divergence (MMD) was designed for local pattern aggregation. The proposed D-WFA was used to train a CNN model to overcome the drawback of the FedAvg algorithm, which resembled systems heterogeneity because clients have different contributions to the overall training. Used data were retrieved by testing a Nippon Seiko Kabushiki-gaisha (NSK) 40BNR10 ball bearing in a milling machine fault-diagnosis experiment. Four bearings were studied under four typical health conditions, namely, healthy (H), inner ring fault (IF), outer ring fault (OF), and broken cage fault (CF).

3.2. Discussion

Table 2 represents an overview to better understand the contributions of FL work in FD. More precisely, it describes the used FL algorithms, machine-learning architecture, involved datasets, and the main challenges the authors were concerned with. Therefore, to better analyze the results in Table 2, we followed a similar methodology of drawing pie charts for each particular case. Figure 4 was therefore introduced for clarification.

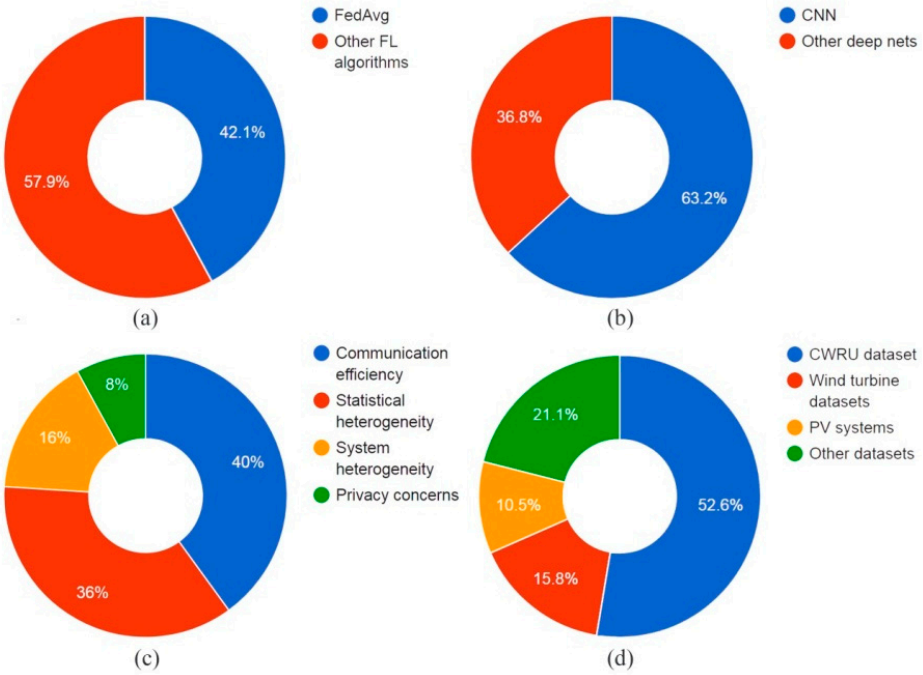

Figure 4. Analysis results of FL works on FD: (**a**) Used FL algorithms for aggregating FD learning machines; (**b**) Adopted architectures for machine learning models; (**c**) Targeted challenges of FL by FD systems; (**d**) Studied datasets for FL and FD works.

Table 2. FL works carried out within the framework of FD.

Year	Ref.	Aggregation Algorithm	Learning Algorithm	Dataset	Solved Challenges
2020	[32]	FedAvg	LSTM	A language-processing-based dataset of a simulated power grid system. The system has cloud computing with 10 edge servers and several terminals.	CoE
2021	[33]	FedAvg	CNN enhanced with data augmentation algorithm	CWRU [34] and Bogie bearing datasets.	StH
	[35]	Kalman filter	Kalman filter	Dataset retrieved from a specific rotating mechanical device (type is not specifically revealed).	CoE
	[36]	Kalman filtering enhanced by a real-time participant-identification method	Kalman filter	Dataset retrieved from a specific rotating mechanical device (type was not specifically revealed). The CWRU dataset [34] was also studied in this work.	Communication efficiency
	[37]	FedAvg	CNN	CWRU [34] and MFD [38].	StH.
	[39]	MBNN	MBNN	CWRU [34] and the gearbox failure diagnostic dataset [40,41].	CoE and SyH
	[42]	FedAvg	SSAE and Siamese networks	Inter-turn short-circuit (ITSC) faults dataset.	Communication efficiency
	[43]	FedAvg with some adaptive-learning features	CNN with a dropout layer trained by MGD	CWRU [34].	CoE
	[44]	FedAvg	Deep adversarial semi-supervised network and a CNN	CWRU [34] and Bogie datasets.	SyH
2022	[45]	Transfer learning	CNN	CWRU [34] and CRACK datasets.	StH
	[46]	FedAvg	Multiscale residual attention network and a custom deep network designed for classification	A 3 MW direct-drive wind turbine [47] and 1.5 MW direct-drive WT [48] fault datasets.	StH
	[49]	FedMM	CNN with adversarial learning features	Gas-insulated switchgear insulation fault-detection dataset.	StH
	[50]	Paillier cryptographic and adaptive control algorithm	CNN	CWRU [34]	PrC, StH, and CoE
	[51]	FedAvg and particle swarm optimization for client weighting	Custom deep network designed for classification	Turns in permanent magnet synchronous motor short-circuit fault-detection dataset.	StH, and CoE
	[52]	ADFL	CNN	Real PV generator consisting of two PV strings connected in parallel with 22 PV modules.	SyH, StH, and CoE
	[53]	FA-FedAvg	CNN	CWRU [34]	CoE
	[54]	Biometric authorization mechanism, gradient noise mechanism, and proportional parameter update strategy	ResNet18 [55]	5 biometric authentication wind farm datasets.	PrC
	[56]	FTL-ASL	CNN	PUD [57], MFPT [58], and CWRU [34] datasets.	StH
	[59]	D-WFA and MMD	CNN	(NSK) 40BNR10 ball bearing dataset.	SyH

Hence, by considering the studied FL algorithms as the main investigation criteria (Figure 4a), we found that the standard FedAvg algorithm was always used (42.1%), while the new variants (57.9%) were used more than FedAvg. FedAvg was generally used when the main challenges addressed were not primarily related to StH or SyH. The reason behind this is that FedAvg considers systems or participant data equally contributing to the training process with the same weights. However, other variants generally add an enhancement

to FedAvg or perhaps use other algorithms rather than FedAvg to update local models while each participant has their own weight in the context of the data and the nature of the system (device). Additionally, for work on CoE, standard FedAvg cannot handle this issue. This is why new features, such as client selection, synchronization, and compression, are needed. In the context of privacy, FedAvg can be used with additional features that ensure FL process immunity, such as encryption.

Figure 4b depicts machine-learning algorithms. Generally, complex deep networks such as generative adversarial networks, LSTM, CNN, and auto-encoders are used in the FD process. However, among all, and since the problem of FD is a classification problem, CNN and its variants, which are very powerful pattern classification tools, dominate, represented by over 63.2% of the algorithms used, while ResNet18 also comes into play. CNN is typically an architecture that follows a non-adaptive training process. Projecting this problem onto the nature of FD data in general and the nature of the training process will be one of its limitations.

When subject to targeted challenges (Figure 4c), StH and SyH are almost equally considered and receive more attention than CoE and PrC. Contrary to what is recommended by the pie chart in Figure 2, privacy receives little attention in contrast to practical requirements. This can be seen as an issue with systems designed for security.

In the context of the used datasets (Figure 4d), bearing datasets are the most investigated, especially CWRU. Little interest has been focused on other datasets. Additionally, these datasets are not specifically designed for FL, and their partitioning does not purely describe an accepted FL process for the work in [46], which perfectly describes a real non-IID situation.

3.3. Challenges

In terms of the previous review results, FL algorithms designed for FD are subject to many challenges, making their investigation in real CM systems questionable under some situations, as indicated below.

Data unavailability: Generally speaking, used datasets in these FD works are meant to be used for offline training for a particular device in the first place and are not specifically designed for FL. It is true they might cope with changes in working conditions, or they could be divided to reach such FL conditions, but this is not enough to simulate a distributed environment where data could perfectly address a real non-IID problem. This limits the generalization capacity of developed models. This data unavailability also does not support CoE and PrC, where many devices with specific resource constraints, communications methods, and protocols can be used in real conditions.

Generalization in case studies: The provided case studies typically dealt with a dataset of industrial systems instead of real industrial processes, including real interactions between different participants and devices with different architectures. This means that FL challenges are not fully addressed, and it is mostly the non-IID cases that are addressed. In this case, the main challenge is to provide a simulation of a real federated network describing fault-detection scenarios to mimic reality in support of the conclusions and not only for the purpose of evaluating the accuracy of the learning model.

The rarity of real failure scenarios: Industrial processes are subject to a preventive maintenance program ensuring a reduction in downtime. The collection of such scenarios is more challenging, especially from a federated network, to provide real conclusions. This leads to many alternative solutions, such as simulations and accelerated aging, which are already observed in pie charts of the investigated datasets.

FL algorithm privacy: The designed algorithms perfectly consider StH and SyH while almost completely ignoring CoE and PrC because it is challenging to develop an algorithm that is secure without suffering from less accuracy. Security features such as encryption lead to model parameter distortion, generating diminished aggregation and prediction performances.

Algorithmic complexity: Complex deep learning architecture, such as CNNs, could result in different drawbacks related to computational costs during training as well as communication. For instance, a huge number of learning parameters resulting from deep-layer architectures makes addressing an EoC with fewer packets and fewer rounds difficult under FL.

Adaptive learning: It is common that an industrial process is subject to massive changes in working conditions due to either the internal or external characteristics of field equipment. Additionally, it is known that adaptive learning is the key feature to successful prediction. However, under FL constraints related to CoE and PrC, it is difficult to reach a higher level of dynamic programming while keeping system performances at the same level.

3.4. Future Prospects

With regard to previous FD challenges under FL networks, the following guidelines should be followed to improve performance.

FD datasets generations for FL networks: To provide a further generalized study on FD in FL networks, emulating real experiments on real industrial processes is required (e.g., massive dynamic data, imbalanced classes, different data partitioning, and different working conditions). This will facilitate data in good agreement with the real conditions of SyS, StH, PrC, and CoE.

Generative modeling and transfer learning: To overcome the lack of faulty patterns in FL networks due to simulation and accelerated aging, generative modeling should receive satisfactory attention when dealing with this type of data. This will help provide samples that are more robust or at least improve the accuracy of predictions of unseen samples.

Addressing SyH and StH in real FL networks: The work provided so far deals with datasets of the same industrial equipment. Therefore, future work should consider heterogeneity by providing different systems belonging to the same FL network to provide further conclusions about FD and FL.

Addressing CoE: The provided work did not account for the different types of sensors (e.g., wired/wireless) and measurements (e.g., sampling rates) in real industrial processes and federated networks. This limits the conclusions obtained to only model accuracy and not FL and FD in smart infrastructures. Therefore, further investigations should consider the CoE issue to provide additional conclusions on FL algorithms and connectivity.

Reducing the complexity of learning algorithms: Focusing on studying simpler effective architectures, such as in [60], is necessary to overcome expansive communication issues.

Focusing on PrC: More works need to be assigned to the PrC of FL algorithms to prevent any attacks under a simple algorithmic architecture.

Addressing adaptive learning: Adaptive learning needs to be explored to deal with the change in data characteristics and address online learning.

4. Conclusions

This paper was dedicated to reviewing recent advances in FD under FL. Since there is no detailed literature survey on the topic as far as we know, we aimed to feature a set of very interesting review/survey papers that fit into the context of FL in general. This step is considered an important phase in the extraction of the main analysis criteria, such as the stakes, the taxonomy, and the most common topics. The information extracted from the review used to analyze the FD works, respectively, has been carried out so far in the form of FL. The main conclusions obtained about FD are that the algorithms designed by FL suffer from a lack of generalization due to the unavailability of data, in addition to a complex architecture making the training process difficult to perform, especially when considering SyH, StH, PrC, and CoE together. To continue to achieve the aims of this work, new steps will be dedicated to the targeting of "prognosis and health management" under FL for more generalization on the study of CM.

Author Contributions: Conceptualization, T.B. (Tarek Berghout); methodology, T.B. (Tarek Berghout); formal analysis, T.B. (Tarek Berghout), M.B., T.B. (Toufik Bentrcia), W.H.L. and Y.A.; investigation, T.B. (Tarek Berghout), M.B., T.B. (Toufik Bentrcia), W.H.L. and Y.A.; writing—original draft preparation, T.B. (Tarek Berghout); writing—review and editing, T.B. (Tarek Berghout), M.B., T.B. (Toufik Bentrcia), W.H.L. and Y.A.; visualization, T.B. (Tarek Berghout), M.B., T.B. (Toufik Bentrcia), W.H.L. and Y.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ramu, S.P.; Boopalan, P.; Pham, Q.V.; Maddikunta, P.K.R.; Huynh-The, T.; Alazab, M.; Nguyen, T.T.; Gadekallu, T.R. Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. *Sustain. Cities Soc.* **2022**, *79*, 103663. [CrossRef]
2. Banabilah, S.; Aloqaily, M.; Alsayed, E.; Malik, N.; Jararweh, Y. Federated learning review: Fundamentals, enabling technologies, and future applications. *Inf. Process. Manag.* **2022**, *59*, 103061. [CrossRef]
3. Aledhari, M.; Razzak, R.; Parizi, R.M.; Saeed, F. Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications. *IEEE Access* **2020**, *8*, 140699–140725. [CrossRef] [PubMed]
4. Berghout, T.; Benbouzid, M. A Systematic Guide for Predicting Remaining Useful Life with Machine Learning. *Electronics* **2022**, *11*, 1125. [CrossRef]
5. Berghout, T.; Benbouzid, M.; Muyeen, S.M. Machine Learning for Cybersecurity in Smart Grids: A Comprehensive Review-based Study on Methods, Solutions, and Prospects. *Int. J. Crit. Infrastruct. Prot.* **2022**, *38*, 100547. [CrossRef]
6. McMahan, B.; Moore, E.; Ramage, D.; Hampson, S.; y Arcas, B.A. Communication-Efficient Learning of Deep Networks from Decentralized Data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS 2017, Fort Lauderdale, FL, USA, 20–22 April 2017; Volume 54, pp. 1273–1282.
7. Alazab, M.; RM, S.P.; M, P.; Maddikunta, P.K.R.; Gadekallu, T.R.; Pham, Q.-V. Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions. *IEEE Trans. Ind. Inform.* **2022**, *18*, 3501–3509. [CrossRef]
8. Agrawal, S.; Sarkar, S.; Aouedi, O.; Yenduri, G.; Piamrat, K.; Alazab, M.; Bhattacharya, S.; Maddikunta, P.K.R.; Gadekallu, T.R. Federated Learning for intrusion detection system: Concepts, challenges and future directions. *Comput. Commun.* **2022**, *195*, 346–361. [CrossRef]
9. Yang, Q.; Liu, Y.; Cheng, Y.; Kang, Y.; Chen, T.; Yu, H. Federated Learning. *Synth. Lect. Artif. Intell. Mach. Learn.* **2020**, *13*, 1–207. [CrossRef]
10. Jatain, D.; Singh, V.; Dahiya, N. A contemplative perspective on federated machine learning: Taxonomy, threats & vulnerability assessment and challenges. *J. King Saud Univ.-Comput. Inf. Sci.* **2021**, *34*, 6681–6698. [CrossRef]
11. Wahab, O.A.; Mourad, A.; Otrok, H.; Taleb, T. Federated Machine Learning: Survey, Multi-Level Classification, Desirable Criteria and Future Directions in Communication and Networking Systems. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1342–1397. [CrossRef]
12. Yang, Q.; Liu, Y.; Chen, T.; Tong, Y. Federated machine learning: Concept and applications. *ACM Trans. Intell. Syst. Technol.* **2019**, *10*, 1–19. [CrossRef]
13. Li, L.; Fan, Y.; Tse, M.; Lin, K.Y. A review of applications in federated learning. *Comput. Ind. Eng.* **2020**, *149*, 106854. [CrossRef]
14. Google Tensorflow Federated Learning. Available online: https://www.tensorflow.org/federated (accessed on 17 August 2022).
15. Webank Federated AI Technology Enabler. Available online: https://github.com/webank (accessed on 17 August 2022).
16. Mothukuri, V.; Parizi, R.M.; Pouriyeh, S.; Huang, Y.; Dehghantanha, A.; Srivastava, G. A survey on security and privacy of federated learning. *Future Gener. Comput. Syst.* **2021**, *115*, 619–640. [CrossRef]
17. Yin, X.; Zhu, Y.; Hu, J. A Comprehensive Survey of Privacy-preserving Federated Learning. *ACM Comput. Surv.* **2021**, *54*, 1–36. [CrossRef]
18. Zhang, C.; Xie, Y.; Bai, H.; Yu, B.; Li, W.; Gao, Y. A survey on federated learning. *Knowl.-Based Syst.* **2021**, *216*, 106775. [CrossRef]
19. Blanco-Justicia, A.; Domingo-Ferrer, J.; Martínez, S.; Sánchez, D.; Flanagan, A.; Tan, K.E. Achieving security and privacy in federated learning systems: Survey, research challenges and future directions. *Eng. Appl. Artif. Intell.* **2021**, *106*, 104468. [CrossRef]
20. Jawadur Rahman, K.M.; Ahmed, F.; Akhter, N.; Hasan, M.; Amin, R.; Aziz, K.E.; Muzahidul Islam, A.K.M.; Mukta, M.S.H.; Najmul Islam, A.K.M. Challenges, Applications and Design Aspects of Federated Learning: A Survey. *IEEE Access* **2021**, *9*, 124682–124700. [CrossRef]
21. Nguyen, D.C.; Ding, M.; Pathirana, P.N.; Seneviratne, A.; Li, J.; Vincent Poor, H. Federated Learning for Internet of Things: A Comprehensive Survey. *IEEE Commun. Surv. Tutor.* **2021**, *23*, 1622–1658. [CrossRef]
22. Briggs, C.; Fan, Z.; Andras, P. A Review of Privacy-Preserving Federated Learning for the Internet-of-Things. In *Studies in Computational Intelligence*; Springer International Publishing: New York, NY, USA, 2021; Volume 965, pp. 21–50. ISBN 9783030706043.
23. Zhu, H.; Xu, J.; Liu, S.; Jin, Y. Federated learning on non-IID data: A survey. *Neurocomputing* **2021**, *465*, 371–390. [CrossRef]

24. Yang, Q. Toward Responsible AI: An Overview of Federated Learning for User-centered Privacy-preserving Computing. *ACM Trans. Interact. Intell. Syst.* **2021**, *11*, 1–22. [CrossRef]
25. Abreha, H.G.; Hayajneh, M.; Serhani, M.A. Federated Learning in Edge Computing: A Systematic Survey. *Sensors* **2022**, *22*, 450. [CrossRef]
26. Ma, X.; Zhu, J.; Lin, Z.; Chen, S.; Qin, Y. A state-of-the-art survey on solving non-IID data in Federated Learning. *Future Gener. Comput. Syst.* **2022**, *135*, 244–258. [CrossRef]
27. Shaheen, M.; Farooq, M.S.; Umer, T.; Kim, B.-S. Applications of Federated Learning; Taxonomy, Challenges, and Research Trends. *Electronics* **2022**, *11*, 670. [CrossRef]
28. Gupta, R.; Alam, T. Survey on Federated-Learning Approaches in Distributed Environment. *Wirel. Pers. Commun.* **2022**, *125*, 1631–1652. [CrossRef]
29. Li, T.; Sahu, A.K.; Zaheer, M.; Sanjabi, M.; Talwalkar, A.; Smith, V. Federated Optimization in Heterogeneous Networks. *Proc. Mach. Learn. Syst.* **2018**, *2*, 429–450.
30. Reisizadeh, A.; Mokhtari, A.; Hassani, H.; Jadbabaie, A.; Pedarsani, R. FedPAQ: A Communication-Efficient Federated Learning Method with Periodic Averaging and Quantization. *arXiv* **2019**, arXiv:1909.13014. [CrossRef]
31. So, J.; Guler, B.; Avestimehr, A.S. Turbo-Aggregate: Breaking the Quadratic Aggregation Barrier in Secure Federated Learning. *IEEE J. Sel. Areas Inf. Theory* **2021**, *2*, 479–489. [CrossRef]
32. Hou, S.; Lu, J.; Zhu, E.; Zhang, H.; Ye, A. A Federated Learning-Based Fault Detection Algorithm for Power Terminals. *Math. Probl. Eng.* **2022**, *2022*, 9031701. [CrossRef]
33. Zhang, W.; Li, X.; Ma, H.; Luo, Z.; Li, X. Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. *Knowl.-Based Syst.* **2021**, *213*, 106679. [CrossRef]
34. Bearing Data Center (CRWU) Seeded Fault Test Data. Available online: https://engineering.case.edu/bearingdatacenter (accessed on 22 December 2022).
35. Xue, M.A.; Chenglin, W.E.N. An Asynchronous Quasi-Cloud/Edge/Client Collaborative Federated Learning Mechanism for Fault Diagnosis. *Chin. J. Electron.* **2021**, *30*, 969–977. [CrossRef]
36. Ma, X.; Wen, C.; Wen, T. An Asynchronous and Real-time Update Paradigm of Federated Learning Diagnosisfor Fault. *IEEE Trans. Ind. Inform.* **2021**, *3203*, 8531–8540. [CrossRef]
37. Li, Z.; Li, Z.; Li, Y.; Tao, J.; Mao, Q.; Zhang, X. An intelligent diagnosis method for machine fault based on federated learning. *Appl. Sci.* **2021**, *11*, 12117. [CrossRef]
38. Marins, M.A.; Ribeiro, F.M.L.; Netto, S.L.; da Silva, E.A.B. Improved similarity-based modeling for the classification of rotating-machine failures. *J. Franklin Inst.* **2018**, *355*, 1913–1930. [CrossRef]
39. Wang, Q.; Li, Q.; Wang, K.; Wang, H.; Zeng, P. Efficient federated learning for fault diagnosis in industrial cloud-edge computing. *Computing* **2021**, *103*, 2319–2337. [CrossRef]
40. Gear Fault Data. Available online: https://figshare.com/articles/Gear_Fault_Data/6127874/1 (accessed on 22 December 2022).
41. Cao, P.; Zhang, S.; Tang, J. Preprocessing-Free Gear Fault Diagnosis Using Small Datasets With Deep Convolutional Neural Network-Based Transfer Learning. *IEEE Access* **2018**, *6*, 26241–26253. [CrossRef]
42. Zhang, J.; Wang, Y.; Zhu, K.; Zhang, Y.; Li, Y. Diagnosis of Inter-Turn Short Circuit Faults in Permanent Magnet Synchronous Motors Based on Few-Shot Learning under a Federated Learning Framework. *IEEE Trans. Ind. Inform.* **2021**, *3203*, 8495–8504. [CrossRef]
43. 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. [CrossRef]
44. Zhang, W.; Li, X. Federated Transfer Learning for Intelligent Fault Diagnostics Using Deep Adversarial Networks with Data Privacy. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 430–439. [CrossRef]
45. Zhang, W.; Li, X. Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions. *Struct. Health Monit.* **2022**, *21*, 1329–1344. [CrossRef]
46. Jiang, G.; Fan, W.P.; Li, W.; Wang, L.; He, Q.; Xie, P.; Li, X. DeepFedWT: A federated deep learning framework for fault detection of wind turbines. *Meas. J. Int. Meas. Confed.* **2022**, *199*, 111529. [CrossRef]
47. Leahy, K.; Hu, R.L.; Konstantakopoulos, I.C.; Spanos, C.J.; Agogino, A.M. Diagnosing wind turbine faults using machine learning techniques applied to operational data. In Proceedings of the 2016 IEEE International Conference on Prognostics and Health Management (ICPHM), Ottawa, ON, Canada, 20–22 June 2016; pp. 1–8.
48. Yuan, B.; Wang, C.; Luo, C.; Jiang, F.; Long, M.; Yu, P.S.; Liu, Y. WaveletAE: A Wavelet-enhanced Autoencoder for Wind Turbine Blade Icing Detection. *arXiv* **2019**, arXiv:1902.05625. [CrossRef]
49. Wang, Y.; Yan, J.; Yang, Z.; Dai, Y.; Wang, J.; Geng, Y. A Novel Federated Transfer Learning Framework for Intelligent Diagnosis of Insulation Defects in Gas-Insulated Switchgear. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3517711. [CrossRef]
50. 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*, 6844–6854. [CrossRef]
51. 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*, 16835–16849. [CrossRef]
52. Liu, Q.; Yang, B.; Wang, Z.; Zhu, D.; Wang, X.; Ma, K.; Guan, X. Asynchronous Decentralized Federated Learning for Collaborative Fault Diagnosis of PV Stations. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 1680–1696. [CrossRef]

53. Geng, D.Q.; He, H.W.; Lan, X.C.; Liu, C. Bearing fault diagnosis based on improved federated learning algorithm. *Computing* **2022**, *104*, 1–19. [CrossRef]
54. Lu, S.; Gao, Z.; Xu, Q.; Jiang, C.; Zhang, A.; Wang, X. Class-Imbalance Privacy-Preserving Federated Learning for Decentralized Fault Diagnosis With Biometric Authentication. *IEEE Trans. Ind. Inform.* **2022**, *18*, 9101–9111. [CrossRef]
55. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.* **2016**, *2016*, 770–778. [CrossRef]
56. Yang, W.; Chen, J.; Chen, Z.; Liao, Y.; Li, W. Federated Transfer Learning for Bearing Fault Diagnosis Based on Averaging Shared Layers. In Proceedings of the 2021 Global Reliability and Prognostics and Health Management (PHM-Nanjing), Nanjing, China, 15–17 October 2021; pp. 1–7. [CrossRef]
57. Paderborn University. Available online: https://mb.uni-paderborn.de/en/kat/main-research/datacenter/bearing-datacenter/data-sets-and-download (accessed on 22 December 2022).
58. Eric, B. Condition Based Maintenance Fault Database for Testing of Diagnostic and Prognostics Algorithms. Available online: https://www.mfpt.org/fault-data-sets/ (accessed on 22 December 2022).
59. Chen, J.; Li, J.; Huang, R.; Yue, K.; Chen, Z.; Li, W. Federated Transfer Learning for Bearing Fault Diagnosis With Discrepancy-Based Weighted Federated Averaging. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3514911. [CrossRef]
60. Berghout, T.; Bentrcia, T.; Ferrag, M.A.; Benbouzid, M. A Heterogeneous Federated Transfer Learning Approach with Extreme Aggregation and Speed. *Mathematics* **2022**, *10*, 3528. [CrossRef]