Analysis of Statistical Data Heterogeneity in Federated Fault Identification

Zahra Taghiyarrenani, Sławomir Nowaczyk, and Sepideh Pashami

Center for Applied Intelligence Systems Research, Halmstad University, Halmstad, Sweden
{*firstname.lastname*}*@hh.se*

ABSTRACT

Federated Learning (FL) is a setting where different clients collaboratively train a Machine Learning model in a privacy-preserving manner, i.e., without the requirement to share data. Given the importance of security and privacy in real-world applications, FL is gaining popularity in many areas, including predictive maintenance. For example, it allows independent companies to construct a model collaboratively. However, since different companies operate in different environments, their working conditions may differ, resulting in heterogeneity among their data distributions. This paper considers the fault identification problem and simulates different scenarios of data heterogeneity. Such a setting remains challenging for popular FL algorithms, and thus we demonstrate the considerations to be taken into account when designing federated predictive maintenance solutions.

1. INTRODUCTION

The emergence of Machine Learning (ML) for prognostics and health management applications, including fault identification, has significantly benefited the field. However, in order to construct predictive models, ML methods require access to sufficient training data. Considering that different companies may own data, even if sharing it would benefit all of them, privacy-preserving and security-related problems may prevent data owners from doing that.

Using Federated Learning (FL), on the other hand, different data owners collaborate to train a model while preserving privacy, i.e., without sharing data. As an example, we may consider a fleet of diverse vehicles owned and operated by different actors. Each vehicle's data is used for training a local model through edge processing. These local models are then transferred to a central server, possibly managed by the original equipment manufacturer (OEM), to construct the global model. The global model can then generalize to all these vehicles and all usage conditions, likely achieving performance that would be impossible to obtain by any of the actors alone. However, many challenges are involved in finding the right solution within the federated learning setting. In this work, we primarily consider the clients' situations, and how they differ. Our setup assumes that all clients are connected to a central server and there is no direct connection between the clients. During the communication round, the clients share only their models (never their data), and only with the server (never with other clients). All clients operate with the same set of features and the label set. On the other hand, we assume that the data distribution varies across clients. Statistical data heterogeneity has always been noted as one of the most challenging aspects of FL, as it adversely affects performance. However, it is crucial to consider this challenge when solving predictive maintenance (PM) problems since different operating and environmental conditions are inevitable in real-world industries and inherently introduce such statistical heterogeneity.

This paper analyzes different variants of statistical data heterogeneity and illustrates the considerations for designing an FL method. We explore different situations and provide results using FedAvg (McMahan et al., 2017), the most popular FL method, to support our analysis. Note that the purpose of this report is not to evaluate FedAvg specifically but rather to highlight the challenges that must be considered when designing effective FL solutions for health monitoring.

2. RELATED WORKS

The nature of FL requires several factors to be considered that are beyond conventional ML. FL has been studied from a variety of perspectives in various surveys and review papers. Several different setups and their corresponding challenges are discussed in (Yang et al., 2019), including vertical, horizontal, and federated transfer learning. According to that taxonomy, the setup considered here is horizontal FL. Zhao et al., 2023, discuss FL from the communication efficiency point of view. An overview of FL applications can be found in C. Zhang et al., 2021. There are also surveys that are specific to certain fields; (Pandya et al., 2023) discuss FL methods and challenges in the context of smart cities, while

(Nguyen et al., 2021; Khan et al., 2021) provide a taxonomy of FL for IoT services and applications.

Most of the surveys listed above mention the problem of heterogeneous data as a challenge that exists for many applications. In the context of PM, several papers solve the problem of fault identification, including (W. Zhang, Li, Ma, Luo, & Li, 2021). Even though these papers focus on designing new methods, they examine three scenarios for evaluation: IID distributed, Non-IID-Class, and Non-IID-Domain; this is, however, different from our work, where we analyze possible data heterogeneous scenarios and their effects on the FL.

3. Federated Learning

This section defines the Federated Learning setup considered in this paper.

We assume a fixed number of K clients $\{c_i | i = 1, ..., K\}$. Given X and Y as input and output spaces, respectively, dataset $D_k = \{(x_i, y_i) | i = 1, ..., n_k\}$ is the local data of client c_k, such that $D_k \sim P_k(x, y)$, a local probability distribution. Each local probability distribution originates from some *global* probability distribution, denoted $Q_l(x, y)$.

The reason for this distinction between P and Q is the potential for conflicts within the data. For example, two clients may have different data distributions $P_i(x, y) \neq P_j(x, y)$ due to diverse sampling. Still, if both originate from the same $Q(x, y)$, a consistent conditional probability distribution $P_i(y|x) = P_j(y|x)$ is guaranteed to exist, and thus a compatible model can be trained across those clients. On the other hand, if P_i and P_j originate from two different Q_1 and Q_2, such consistency is not assured.

Every client first constructs a *local model* using its own dataset. ***Federated learning*** aims to construct a *global model* that generalizes to each of the clients' distribution. The most straightforward method to construct the global model is averaging the weights of the local models, called FedAvg (McMahan et al., 2017). FedAvg's federated communication rounds are as follows: first, the server initializes a model and sends it to the clients. The clients train the model and send it back to the server. By averaging the local models, the server constructs a global model. This process will continue until the global model converges.

4. Data Heterogeneity in FL

We discuss and review four different heterogeneity scenarios considering local and global distributions $P(x, y)$ and $Q(x, y)$. For each scenario, we present the corresponding situation that may happen for the fault identification problems so that in the experiment section, we can simulate such a situation and analyze the results.

4.1. Unique global and Uniform local

This first scenario meets the following criteria:

- $\forall c_i, c_j \in clients : P_i(x, y) \sim P_j(x, y)$
- $\forall c_i \in clients : P_i(x, y) \sim Q(x, y)$

In this scenario, all clients follow a single global distribution. Moreover, all the clients follow the same local distribution. An example is when multiple companies use the same machine that operates only under one load condition; no heterogeneity is observed. If every client has access to enough training samples, they can construct a model which perfectly generalizes to unseen samples.

4.2. Unique global and Heterogeneously distributed

This scenario meets the following criteria:

- $\exists c_i, c_j \in clients : P_i(x, y) \neq P_j(x, y)$
- $\forall c_i \in clients : P_i(x, y) \subsetneq Q(x, y)$

Similar to the previous case, there is only one unique global distribution, and all clients are part of it. Local distributions, however, differ from one another.

Sometimes clients' working conditions are the same, but the clients do not have enough data to develop a model that can be generalized well. In other words, every local dataset forms part of the global dataset but it does not represent the *whole* global distribution. As a result, although there is no direct conflict between the samples of the different clients, their models differ from one another; we describe this as Empirical Concept Shift. For example, different companies use the same machine under one load condition but in distinct environmental conditions. Certain faults are, therefore, more likely to occur in some of them, leading to data imbalance across clients.

4.3. Mixed global and Uniform local

This scenario meets the following criteria:

- $\exists c_i, c_j \in clients : P_i(x, y) \neq P_j(x, y)$
- $P_i(x, y) \sim Q_1(x, y)$ and $P_j(x, y) \sim Q_2(x, y)$
- $Q_1(x, y) \neq Q_2(x, y)$

In this case, client data originates from more than one global distribution. In particular, some overlap (or conflicts) may exist between different clients' classes. At the same time, there is no conflict within the local data of any client, as each individual $P_i(x, y)$ follows only one global distribution.

An example would be companies with the same type of machine, but these machines in one company work under different load conditions than in another. Thus, the data that is perfectly normal in one of them might indicate a fault in the other.

4.4. Mixed global and Heterogeneous local

This final scenario meets the following criteria:

- $\exists c_i, c_j \in clients : P_i(x, y) \neq P_j(x, y)$,

- $\exists c_i \in clients : P_i(x,y) \subseteq \bigcup_{l=1}^{L} Q_l(x,y)$

Now not only does there exist heterogeneity between clients, but some clients also encounter heterogeneity within their local data. An example of this scenario would be companies where not only the loading conditions across the companies are different, but also some companies have the equipment working under more than one loading condition. This results in conflict between the samples even inside the local models. Figure 1 summarizes the abovementioned four scenarios. This figure also shows the solutions needed to improve the performance of FL methods from a Domain Adaptation point of view (Blitzer et al., 2007). In the first scenario, no adaptation is necessary since all clients follow the same distribution. The second and third scenarios require inter-client client adaptation since the samples' distribution is different. In the fourth scenario, in addition to inter-client adaptation, intra-client adaptation is also in need since some clients have conflicts within their own samples.

5. Experiments

5.1. Experimental Settings

We use a dataset (Elly Treml et al., 2020) containing measurements of currents, vibrations, and three-phase voltages collected from various locations on a three-phase induction motor. The dataset includes cases of one to four broken rotor bars (BRB), in addition to the healthy operation of the motor. In addition, there are different loading conditions based on the level of mechanical torque. In (Taghiyarrenani & Berenji, 2022), the authors provide an analysis of the differences between the loads, their effects on detecting broken rotor bars, and the robustness of both vibration and current signals. We follow the mentioned paper for the experimental setup and use four out of eight available load levels, including 12.5%, 50%, 62.5%, and 100% of nominal load. According to (Taghiyarrenani & Berenji, 2022), the detection of BRBs is more challenging with current signals. Accordingly, we use only current signals and apply the same pre-processing steps as in the mentioned paper. After splitting the original time domain signals into windows 6667 points in length, we apply FFT to obtain frequency domain records. Min/Max scaling is used for normalization purposes. The classification is accomplished using an MLP network with two hidden layers with sizes of 3333 and 1111 and tanh activation functions.

We simulate different FL scenarios, involving eight clients in each; however, the way in which the data is divided among the clients varies as follows:

Unique global and uniform local. In this scenario, we consider all clients to be working under the same conditions. Therefore, we randomly select the nominal load of 0.5 Nm and then proceeded with the following steps: 1) We separate 25% of all samples as global test samples and the remaining samples as training samples; 2) we divide training samples between the clients uniformly. Thus, we use the same test samples for all clients.

Unique global and Heterogeneously distributed. Similarly to the previous scenario, we select the 0.5 Nm load. This time, however, we select only 20% of training samples and distribute them between 8 clients as local training samples. Therefore, none of the clients has enough training data to build a good model alone.

Mixed global and Uniform local. We simulate this scenario by randomly assigning one load for each client. For the eight clients, the selected loading conditions are, in order, $\{0.5\}$,$\{2.0\}$,$\{2.5\}$,$\{4.0\}$,$\{0.5\}$,$\{2.0\}$,$\{2.5\}$,$\{4.0\}$ Nm.

Mixed global and Heterogeneous local. We simulate this scenario by using more than one load for some clients. More precisely, the selected loading conditions for the eight clients are $\{0.5, 2.0\}$, $\{4.0, 2.0\}$, $\{0.5, 2.0, 2.5\}$, $\{0.5, 2.0, 2.5, 4.0\}$, $\{0.5\}$, $\{2.0\}$, $\{2.5\}$, and $\{4.0\}$ Nm, respectively.

In all cases, we present the following results:

Centralized model. This is a baseline where, unlike FL, all clients freely share their data, and a single model is trained

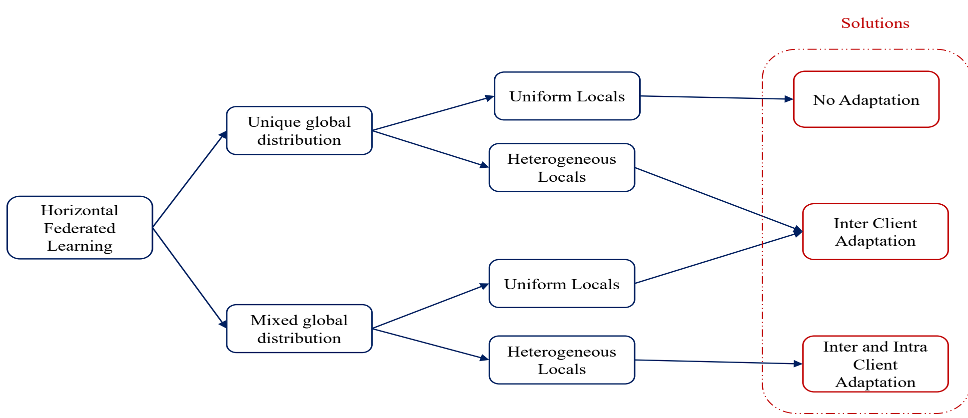

Figure 1. Different categories and solutions considering the distribution of samples.

centrally. For the sake of comparison, we use the same training data as distributed between the clients.

Global model. This is the model trained through Federated Learning. After each communication round, we calculate the performance of the global model.

Clients' models. We show the clients' individual results after every single local epoch in two different setups: with FL (W_FL) and without FL (W/o_FL). In the (W/o_FL) setup, every client trains a model using its own data without participating in federated learning. In the (W_FL) setup, after a number of local training epochs, the clients send the models to the server, and the server calculates the global model and sends it back to the clients. Clients replace their previous models with this new model and continue training using local data. This way, we can examine the effect of FL on the clients.

5.2. Results

We show all results in terms of accuracy on the global test set in Figures 2 to 6. The x-axis shows the local training epochs in all the figures, and the y-axis shows accuracy. The results of the centralized model are in gray, the global model is in black, and the individual clients are in various colors. We also define the Federate Step (FS) as the number of local epochs before the FL communication round.

First, we perform experiments under FS = 50, shown in figures 2 to 5. Figures on the left show all clients' results, while the ones on the right show a random selection of one or two clients. The latter ones are generally more readable and facilitate more in-depth analysis. The goal of the former is to give a high-level overview of how the global model is performing in relation to all clients. In all cases, the centralized model converges rapidly; however, it should be noted that it has access to eight times as many samples as every local model during each epoch, which does not preserve clients' privacy.

Figure 2 corresponds to the first scenario, and we can see that all client models converge completely to the same performance as the global model. The most important finding, shown in the right-hand figure, is that clients will converge to the same level anyway, even without participating in FL. In summary, FL does not provide considerable benefits if different clients have ample samples generated in identical conditions (statistical homogeneity).

Figure 3 shows the results of the second scenario. On the left, one can see how FL benefits the clients since the global model is much better than some local models. After around the fifth communication round, all the client models are upper-bound with the global model. This behavior is consistent with ensemble learning, where an ensemble of weak classifiers performs better than the individual ones. Due to the heterogeneity of clients in this scenario, the diversity between local models is high, which benefits the global model as an ensemble. Furthermore, since there is no conflict between the samples on a global scale, the ensemble will converge to the high-quality centralized model. The results of a client not participating in FL, shown as a blue dashed line in the right subfigure, clearly demonstrate that FL is crucial. In general, this scenario is the best-case setup for FL.

A more in-depth look at the evolution of clients' performance over time reveals another highly significant insight; *Catastrophic Forgetting in Federated Learning*. The phenomenon of catastrophic forgetting is well known in the context of continual learning (De Lange et al., 2021). It means when retraining a trained model with data from a new task, the model starts to forget information about the previous tasks (the one that model was originally trained with).

Figure 3 shows a large drop in the performance of all the local models immediately after the FL communication round. The global model that is replacing a local model has knowledge about other clients, which are statistically different. When a client retrains this model locally, the knowledge from other clients tends to disappear, and consequently, the overall performance of the model drops. This is an important challenge to address in the future in order to increase the efficiency of FL in practical settings.

Figure 4 shows the results of the third scenario. The global model achieves the same performance level as the centralized model. However, removing conflicts between clients could improve the results further, for example, by using Domain Adaptation. It is interesting to note that some of the local models participating in FL outperform the global model, which may be due to the complexity of their inherent data.

Figure 5 shows the results of the fourth scenario. Even though FL is helpful in this scenario, the performance of the global model does not reach the level of the centralized model. This is due to conflicts within the clients, which cause the local models to perform poorly, ultimately leading to a weaker global model. Hence, resolving these internal conflicts is crucial for enhancing the performance of Federated Learning.

Finally, we examine how FS affects the performance of the global model. Figure 6 illustrates the performance, for the second and third scenarios, under three different communication frequencies: every 10, 50, and 100 local epochs. Different FS values have different effects on the models in these two scenarios. Therefore, when designing FL systems, one needs to consider the cost and budget for federated communication and the resources per client. For example, in the case of identical clients, a lower FS leads to faster convergence, which is advantageous when federated communication is less costly than local training.

6. Conclusion

In this paper, we examine the impact of statistical data heterogeneity on Federated Learning (FL) in fault identification applications. Four distinct scenarios were outlined:

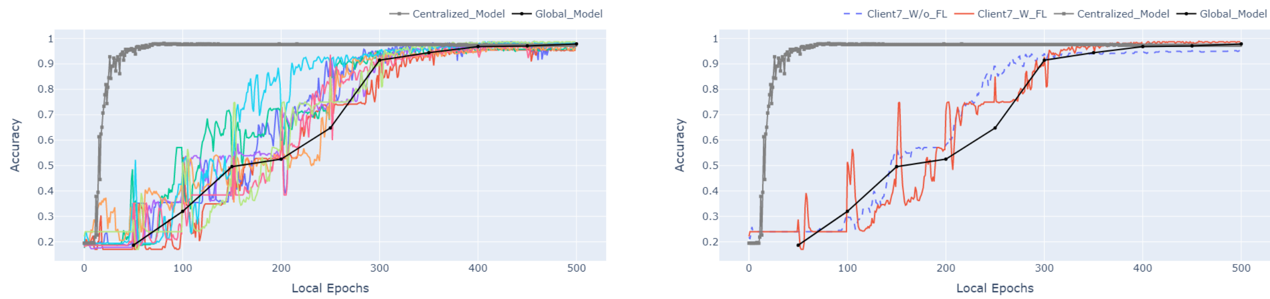

Figure 2. Results of *Unique global and Uniform local* scenario (left: all clients, right: randomly selected client(s)).

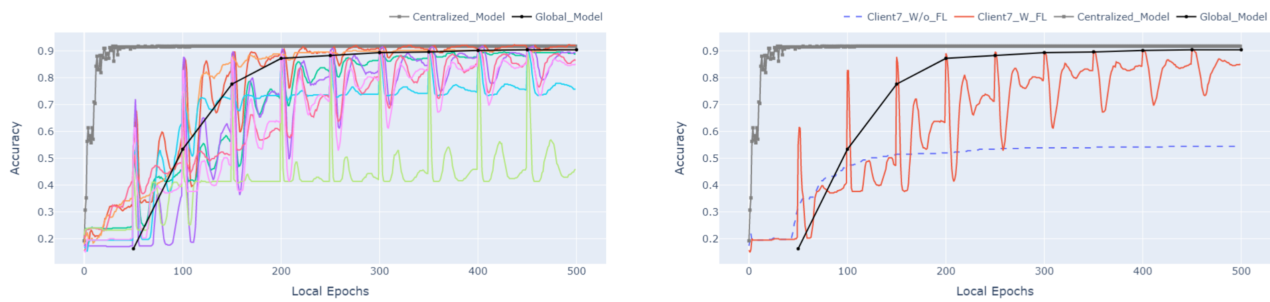

Figure 3. Results of *Unique global and Heterogeneously distributed* scenario (left: all clients, right: randomly selected client(s)).

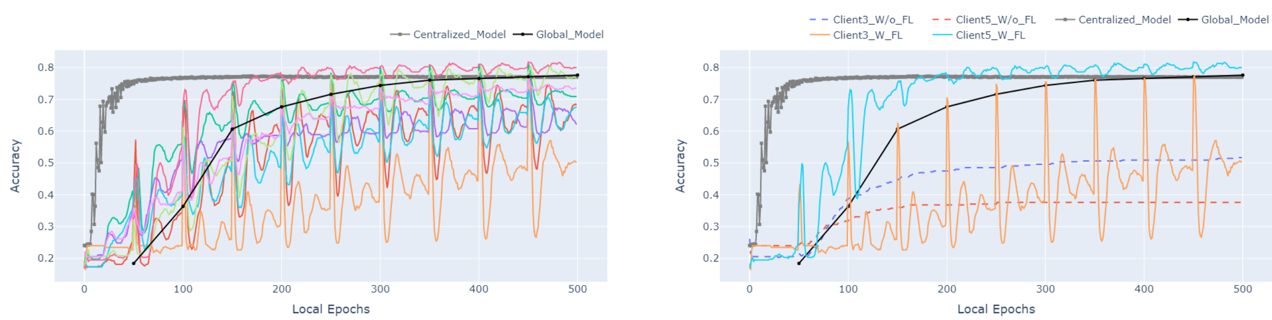

Figure 4. Results of *Mixed global and Uniform local* scenario (left: all clients, right: randomly selected client(s)).

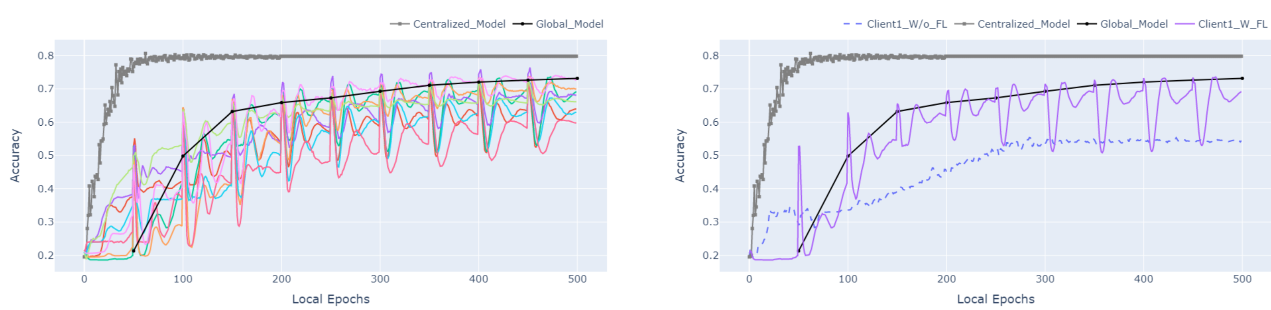

Figure 5. Results of *Mixed global and Heterogeneous local* scenario (left: all clients, right: randomly selected client(s)).

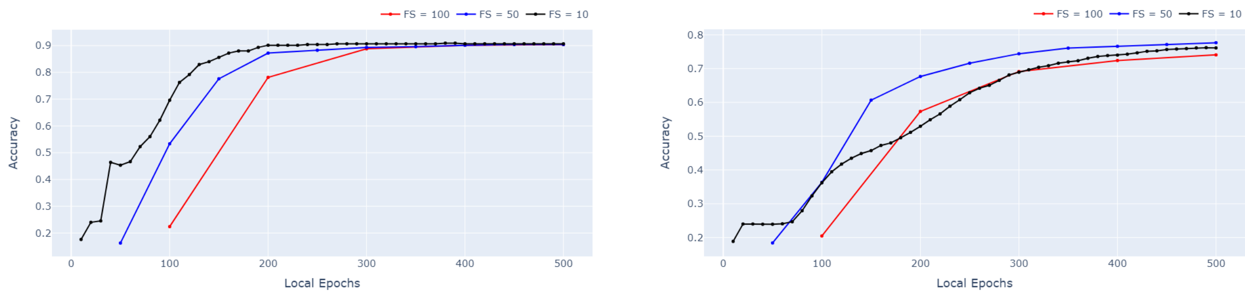

Figure 6. The effect of the FS (Federated step) on the results of the global model; *Unique global and Heterogeneously distributed* scenario on the left and *Unique global and Heterogeneous local* scenario on the right.

1) Uniform distribution across all clients, 2) Each client has data from a section of a single global distribution, resulting in varied local yet consistent global distribution, 3) Divergent global distributions across different clients, causing potential conflicts in globally trained models, and 4) Heterogeneity within the local data of some clients. A practical evaluation of these four scenarios led to the following conclusions: FL may not yield significant improvements in detection accuracy when different clients have ample samples generated under identical conditions. When client data varies, FL proves beneficial. We identified the occurrence of Catastrophic Forgetting in many scenarios; efforts to minimize this phenomenon would be beneficial. Reducing conflicts, whether they occur between clients or within individual clients, is also likely to improve the quality of the results. Finally, we explored the trade-offs between FS (Federated Step) and performance, providing guidance on practical FL setup.

Acknowledgment

This research has been funded by the Vinnova Strategic Vehicle Research and Innovation programme, Sweden.

References

Blitzer, J., Crammer, K., Kulesza, A., Pereira, F., & Wortman, J. (2007). Learning bounds for domain adaptation. *Advances in neural information processing systems*, *20*.

De Lange, M., Aljundi, R., Masana, M., Parisot, S., Jia, X., Leonardis, A., ... Tuytelaars, T. (2021). A continual learning survey: Defying forgetting in classification tasks. *IEEE transactions on pattern analysis and machine intelligence*, *44*(7), 3366–3385.

Elly Treml, A., Andrade Flauzino, R., Suetake, M., & Ravazzoli Maciejewski, N. A. (2020). *Experimental database for detecting and diagnosing rotor broken bar in a three-phase induction motor.* IEEE Dataport.

Khan, L. U., Saad, W., Han, Z., Hossain, E., & Hong, C. S. (2021). Federated learning for internet of things: Recent advances, taxonomy, and open challenges. *IEEE Communic. Surveys & Tutorials*, *23*(3), 1759–1799.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics* (pp. 1273–1282).

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658.

Pandya, S., Srivastava, G., Jhaveri, R., Babu, M. R., Bhattacharya, S., Maddikunta, P. K. R., ... Gadekallu, T. R. (2023). Federated learning for smart cities: A comprehensive survey. *Sustainable Energy Technologies and Assessments*, *55*, 102987.

Taghiyarrenani, Z., & Berenji, A. (2022). An analysis of vibrations and currents for broken rotor bar detection in three-phase induction motors. In *Phm society european conference* (Vol. 7, pp. 43–48).

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Trans. on Int. Systems and Technology (TIST)*, *10*(2), 1–19.

Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., & Gao, Y. (2021). A survey on federated learning. *Knowledge-Based Systems*, *216*, 106775.

Zhang, W., Li, X., Ma, H., Luo, Z., & Li, X. (2021). Federated learning for machinery fault diagnosis with dynamic validation and self-supervision. *Knowledge-Based Systems*, *213*, 106679.

Zhao, Z., Mao, Y., Liu, Y., Song, L., Ouyang, Y., Chen, X., & Ding, W. (2023). Towards efficient communications in federated learning: A contemporary survey. *Journal of the Franklin Institute*.