2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT) | 978-1-6654-7491-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/DICCT56244.2023.10110075

Federated learning: Applications, Security hazards and Defense measures

Sonam Tyagi
School of Computing Graphic Era Hill University Haldwani, India sonamtyagi@gehu.ac.in

Ishwari Singh Rajput *School of Computing Graphic Era Hill University* Haldwani, India ishwarirajput@gehu.ac.in

Richa Pandey
School of Computing Graphic Era Hill University Haldwani, India richapandey@gehu.ac.in

***Abstract*—Federated learning(FL), a cutting-edge method of distributed learning, enables multiple users to share training results while maintaining the privacy of their personal data. Collecting data from different data owners for making machine learning predictions becomes increasingly challenging as data security becomes more of a priority. Federated learning protects user's privacy in addition to increase the training data while overcoming the challenges faced by machine learning and deep learning models. Since the data privacy and security is a world-wide concern, the concept of federated learning is increasing day by day from theoretical to practical level. This review paper involves the overview of the federated learning framework, its types, different applications, several types of attacks and defense mechanism.**

***Index Terms*—Federated learning, Horizontal FL, Vertical FL, Federated TL, Poisoning attack**

I. Introduction

Data security has emerged as a top priority in this age of big data, when nearly all data is transferred across networks. Despite the phenomenal advancement of network data security protection techniques, confidentiality and critical information leakage problems still persist. As computational power increases, machine learning has emerged as the go-to method for analysing and processing large amounts of data, opening up new fields of application [1]. While advancements in machine learning is being made, two issues have emerged that must be addressed. Data security is difficult to secure, and privacy protection is getting more and more crucial. Data sharing, however, is less common in the big data era; as a result of the new trend of sharing data and the increased emphasis on data protection among organisations to prevent data leakage [2]. Unfortunately, data silos have emerged as a result of difficulties in facilitating the exchange of data used for machine learning's training purposes. Google introduced the concept of federated learning with the express aim of breaking the phenomena of data silos [3]. With information security, terminal data privacy, and personal data privacy as its foundations, its design prioritises the effective interchange of large data and the execution of machine learning among numerous users [4]. In addition to neural networks, federated learning can make use of other popular machine learning techniques, such as random forest. In addition to strengthening the safety of enterprise-to-enterprise data exchanges and the precision of training models as a result of such exchanges, the model also successfully addresses the issue of privacy protection while sharing data across organisations [5]. As federated learning has grown, the model's usefulness has spread to a variety of new contexts. Since the end of 2020, major data breaches have occurred on the internet, prompting regulators to probe the majority of large data corporations. There was a data breach affecting 538 million Weibo users in March of 2020. The exposed data included user IDs and phone numbers, basic account information like usernames and avatars, as well as geolocation data and the number of followers each user had. Forty-five million medical pictures from around the world were leaked online on December 16, 2020. Data leaks from big data corporations have put the entire industry under the spotlight, both at home and around the world. The examination of these companies shows that national legislation and regulation are getting stronger, but it also revealed the worldwide scale of data leaks and customer privacy breaches. Clearly, this indicates that diverse businesses should pay attention to the quality of data privacy protection. The Baidu Index reveals that between September 2019 and April 2021, the search index of federated learning keyword drastically increased, indicating that more people are becoming aware of federated learning. How to acquire data, what to do with it, and data security management are all significant concerns in the perspective of modern data privacy protection [6]. With the rise of federated learning, data's true potential can finally be realised [7]. It solves the problems of data privacy and sharing that hamper conventional machine learning and deep learning to a far greater extent than its predecessors do [8]. Since not enough information is collected, the resulting models are inaccurate. This presents a useful challenge for the state-of-the-art AI algorithms that can be used to secure sensitive data.

The key objectives of our study are listed as follows:

1) To present an overview of the current state-of-the-art federated learning concepts.
2) To give a deep insight into the categories of federated learning.
3) Identify the applications of federated learning in various domains.

4) To analyze the Security hazards and Defense measures in federated learning

This article has the following organisational structure. Concepts of federated learning is presented in Section II. The various categories of federated learning is covered in Section III. Applications of federated learning in various domains are discussed in Section IV. Section V mentions various security hazards and their defense measures. The final discussion is presented in Section VI.

II. What is federated learning?

Federated learning objective is to train machine learning algorithm across a number of decentralised edge devices that are holding local data samples without exchanging them. Since the training data is kept locally with members during the federated learning process, this method can not only realise the sharing of each member's training data but also ensure the safeguarding of each member's privacy [9]. The fundamental process of federated learning as shown in Figure 1, consists of following steps:

1) The participant downloads the global model α_{t-1} from the server
2) Participant n trains local data to get local model $\alpha_{t,n}$
3) Participants upload local model updates to the central server
4) After the server receives the data from all participants, it performs a weighted aggregation operation to obtain the global model α_t
   Where, the representations are as follows:
   $\alpha_{t,n}$=local model update of the t^{th} round of communication of the n^{th} participant
   α_t=global model update of the t^{th} round of communication.

The features of the federated learning technology are as below:

1) Data used in federated learning are never transmitted in plain text and are always stored locally on clients, with only model update information requiring interaction with a centralised server.
2) As part of federated learning, everyone who contributes data to the model's α training will have access to the resulting model.
3) Ultimately, a federated learning model's accuracy is comparable to a centrally located machine learning model.
4) The accuracy of the global model increases as the standard of training data used by the federated learning users increases.

III. Types of federated learning

Federated learning has proven to have great potential in privacy-sensitive settings, such as the banking sector, manufacturing, and other areas of data perception. Based on distribution of data, federated learning can be classified as (Figure 2):

- Horizontal federated learning
- Vertical federated learning
- Federated transfer learning

A. Horizontal federated learning

Integration of data sets is the fundamental component of horizontal federated learning. High degrees of similarity between participant data and user data indicate that this is the domain of interest. The data that can be used for collaborative model training is the subset of data where the attributes of the data are shared by both participants [10], despite the fact that the consumers are distinct(Figure 3). In terms of the data, the range of possible situations for horizontal federated learning applications is greater.

B. Vertical federated learning

In vertical federated learning, different parties with multiple attributes about the same set of users can train machine learning models together without sharing either their raw data or their model parameters [11]. When there is more overlap between users and less overlap between data features, the data that can be utilized for joint modelling training is the subset of the data that is not identical in context of data attributes for the same users on both sides(Figure 4).

C. Federated transfer learning

In most circumstances, data does not occupy the same sample area or feature space as in the scenarios of horizontal federated learning and vertical federated learning. Thus, the primary issue here is the scarcity and low quality of data labels [12]. Federated transfer learning is useful when there is little resemblance between the features and samples of the participants(Figure 5). The fundamental concept of federated transfer learning is that various participants have unique attributes.

Advantages of federated learning over traditional learning

Because of its numerous benefits, federated learning is quickly replacing conventional machine learning. Some advantages are as below:

1) Privacy protection of users: To protect users' privacy, the data in federated learning is retained locally, and individual users' data is not shared [13].
2) To train models on massive datasets: The quality of trained models can be enhanced with access to large amounts of training data. Training a model with federated learning offers more accuracy than with conventional machine learning, while also requiring less hardware and allowing for faster training under massive data loads.

How federated learning model works?

1) Federated learning allows for the remote sharing of data by numerous individuals in order to jointly train a single deep learning model and incrementally enhance it.
2) Each party gets the model from a cloud datacentre, which is often a foundation model that has already been trained.

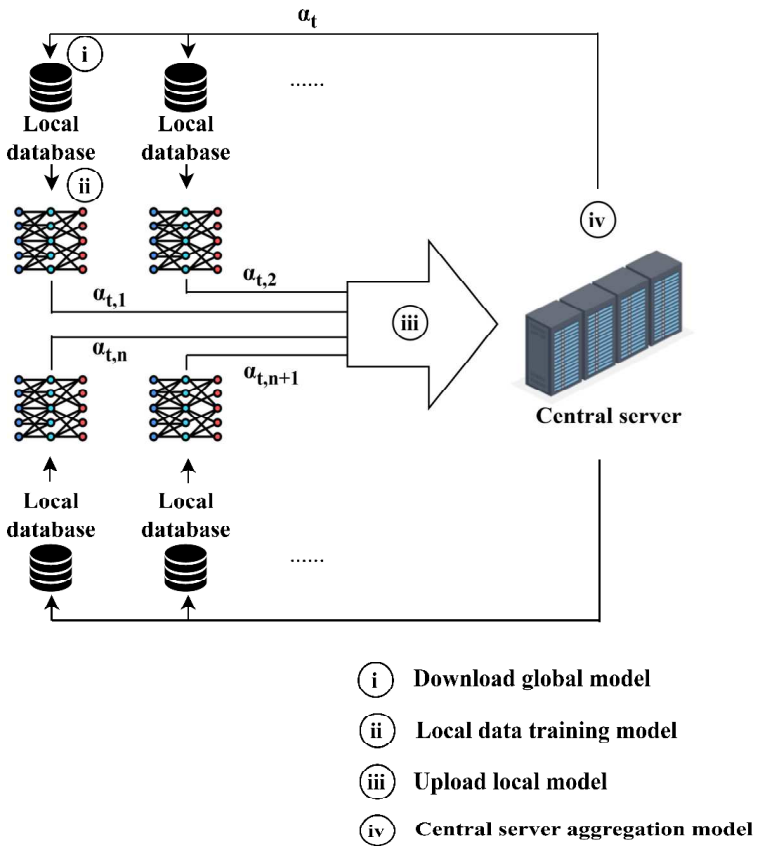

Fig. 1. Federated learning process

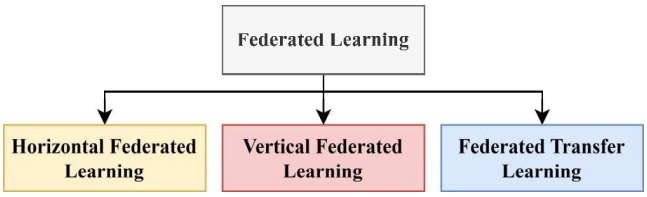

Fig. 2. Classification of federated learning

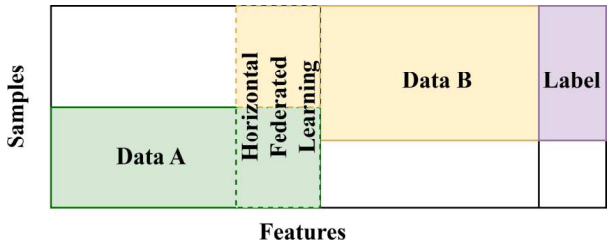

Fig. 3. Horizontal federated learning

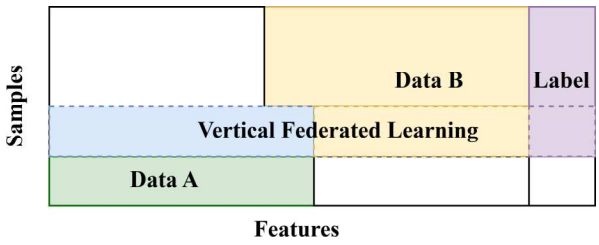

Fig. 4. Vertical federated learning

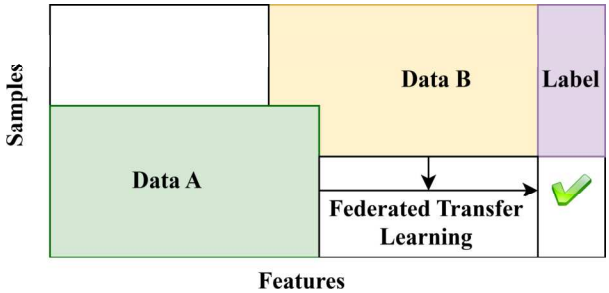

Fig. 5. Federated transfer learning

3) The new configuration of the model is summarised and encrypted once they train it using their personal data.
4) The model changes are transmitted back to the cloud where they are decrypted, averaged, and added to the main model.
5) The collaborative training is carried out until the model

has been fully trained, iteration after iteration.

IV. Application of federated learning

Since its inception, federated learning has been put to a variety of uses, although most of these are still in the experimental phase. The public's knowledge of federated learning has progressed from the theoretical and conceptual to the practical [14]. There are various domains in which federated learning is very useful (Figure 6). Some of them are as below:

A. Intrusion detection

When compared to more conventional deep learning methods, the accuracy of intrusion detection improves when federated learning is used in conjunction with these training procedures. Data privacy is safeguarded, and numerous security problems, including data loss and theft, are avoided as a result [15], [16].

B. Electric power industry

The use of federated learning in the context of the Internet of things with ubiquitous power substantially enhances the processing power issue. Additionally, it satisfies the data distribution needs of federated learning and further enhances data security [17].

C. Financial industry

Data silos is a major problem in the banking sector. All businesses are interested in protecting their customers' personal information and reaping the rewards that come from fully utilising their data. More attention will be paid to data security concerns as a result of increased oversight. The data of diverse financial organisations will have a more noticeable effect if federated learning is appropriately applied to the financial industry to achieve universal applicability [18]. One advantage is that it safeguards the confidentiality of customers' most private financial information. However, the distributed computing of federated learning framework has increased computational power.

D. Communication industry

Due to the particular nature of the communication sector and the massive amounts of user data, privacy protection has evolved into a complex issue. Therefore, implementing federated learning in the communication sector can not only effectively utilise the vast amounts of user data held by operators, but can also guarantee the confidentiality of all user information [19].

E. Healthcare

Because of its potential as a game-changing approach to protecting patient data, federated learning is gaining traction in the healthcare industry [20]. A large amount of patient data may exist at each medical institution, yet that may not be enough to train independent prediction models. Integration of federated learning and illness forecasting is among the good options to reduce the obstacles of analysis across different institutions [21].

V. Security hazards and Défense measures in federated learning

Federated learning has shown great promise and has produced impressive results in a variety of contexts; but, no model is without flaws. When federated learning is integrated with real-world applications, it is inevitably vulnerable to attacks from unethical users and criminals (Figure 7). Although federated learning's benefits are evident and its development is in line with current trends, it still requires extensive testing to ensure its reliability before being applied in practice [22]. There are several threats in federated learning, which are given below:

A. Poisoning attack

When a malicious user manipulates a machine learning model's prediction by attacking on the training dataset, this is called a poisoning attack. In federated learning, poisoning attacks may be carried out in two ways: data poisoning and model poisoning [23]. The term "data poisoning" refers to a scenario in which an attacker corrupts the training set by introducing incorrect labels or biased data, thus lowering the integrity of the samples used to train a model and potentially compromising its trustworthiness or validity. The term "model poisoning" refers to a scenario in which an attacker alters a model without changing the underlying data used to create the model. At the time of global aggregation process, the attacker disturbs the learning process by transmitting erroneous parameters or corrupted models [24].

Defense from poisoning attack: Since in federated learning, poisoning attacks may be carried out in two ways: data poisoning and model poisoning; so the defense may also be considered for these both attacks [25]. The most effective strategy to safeguard data against data poisoning is to guarantee that the data before model training is safe and trustworthy. Second, if it is uncertain if the data being trained is safe and secure, it is essential to ensure that the model is sufficiently stable during training to protect the data's integrity [26]. Since the poisoning party takes part in the poisoning as a normal user, all of its actions will be mistaken for those of a normal user if it is not discovered, making it impossible to detect the poisoned person's data samples. Further, the user of the poison can optimise the model's poisoning attack by adjusting the model's data and parameters adaptively to their own unique poisoning situation.

B. Adversarial attack

An adversarial attack is a deliberate attempt to trick a model into giving false positive results by manipulating its input data. Adversarial sample refers to the input sample created by introducing noise into the original sample [27].

Defense from adversarial attacks: Confrontation training is the most popular strategy for dealing with adversarial attacks. This means that the real dataset and the adversarial sample are combined, trained, and then the resulting model is analysed and refined. While this type of training can improve the

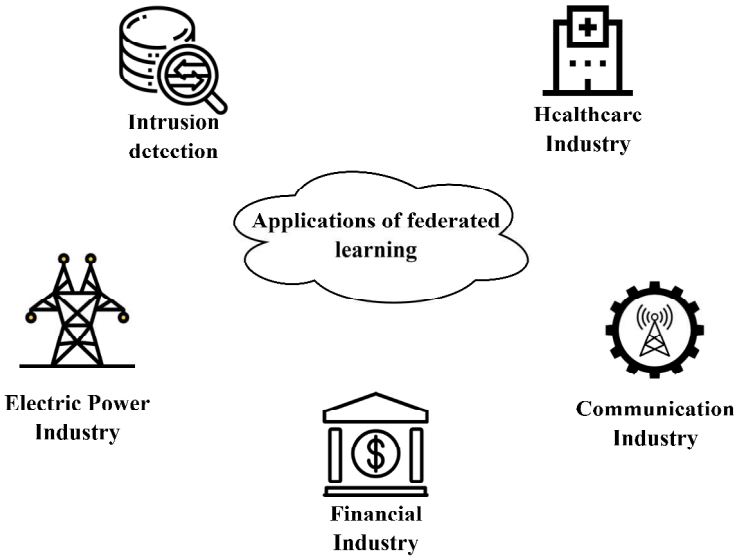

Fig. 6. Applications of federated learning

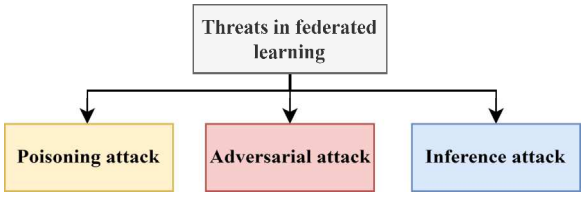

Fig. 7. Threats in federated learning

model's robustness and keep it stable, it also has a clear drawback in that it is susceptible to the trained adversarial sample model but not to any unknown attacks. Data augmentation is another technique to counter adversarial attack [27]. In this case, the original data is randomly altered to improve the model's generalisation capacity, which can be used to counter attacks like image cropping, scaling, steganography, etc.

C. *Inference attack*

As part of the federated learning approach, individuals can train their data locally. Everyone involved works on their own. Indirect access to local information is limited for participants. Privacy and safety can be maintained to some extent. A certain measure of security remains, though, and that is the possibility of privacy leakage. A malicious user, for instance, may utilise the shared parameters to infer the private information of other users and then use that information to create malicious data that would ruin the model [28]. The goal of a forward inference attack is to exploit secrecy of the model by deriving its parameters. The goal of the reverse inference attack is to acquire sensitive user data by inferring it from the training set used to train the model.

Defense from inference attack: Differential privacy, secret sharing mechanisms, homomorphic encryption, and a hybrid defence mechanism are all methods that can be used to resist from inference attacks [29]. Differential Privacy offers statistical protections from the knowledge an adversary might be able to gather from the results of a randomised process. By introducing randomness into the algorithm's calculations, it establishes an absolute maximum on how much a single user can affect the final result. Secret sharing mechanism involves the method in which the confidential information is subdivided and distributed to all of the parties concerned for secure storage. If a certain proportion of secret fragments are not collected, then no information can be acquired, however the hidden information can be altered after collecting those fragments. Another popular and secure type of encryption is homomorphic encryption. It can perform the same action on the ciphertext as it would on the plaintext after decryption without having direct access to the plaintext. Using homomorphic encryption, the server doesn't need to know the user's privacy settings; it just needs to train the ciphertext and aggregate it during aggregation.

VI. Discussion

Using federated learning allows us to leverage multiple data sources and models, enhance system efficiency, in context with the rapidly evolving fields of big data, data security and artificial intelligence. This paper describes a concise introduction to federated learning, explores potential vulnerabilities, and provides solutions to these security concerns in the context of an actual application.

Operating and maintaining a communications network involves a high number of smart devices and results in a massive data output. Direct training on internal data leads to unbalanced data distribution and limited generalisability. If federated learning is extensively used in the communications sector, it has the opportunity to address a number of issues now affecting the sector, as given below:

1) Since data is not dispersed in a consistent manner because different local places include varying amounts of data; hence the accuracy and efficacy of the training models differs.
2) Stored local data cannot be reliably shared due to privacy concerns. It renders typical machine learning training ineffective for generating high-quality models, and it makes interdisciplinary collaboration even more challenging.

Federated learning has the potential to work with a wider variety of machine learning algorithms, though this will ultimately depend on the specifics of each algorithm's communication, encryption, encoding, and so on.

Real-world applications of federated learning typically involve a distributed network of heterogeneous client devices, such as mobile phones and computers located on the network's periphery. Testing and tuning machine learning algorithms operating on mobile apps, such as mobile phones, is more challenging and complex. The problem now is how to get federated algorithms that are both effective and practical. The work that has been done in this area is still needs more exploration by the scholars to conduct deep analysis and study in order to enable it for real-world scenarios. This paper is simply an overview of federated learning, its types, threats and various applications.

In future, it will be extensively applied as a solution to the existing issues in the communication sector. Many people will work together and train under a secure framework to reach a common goal.

REFERENCES

[1] W. Y. B. Lim et al., "Federated Learning in Mobile Edge Networks: A Comprehensive Survey," IEEE Commun. Surv. Tutorials, vol. 22, no. 3, pp. 2031–2063, Jul. 2020, doi: 10.1109/COMST.2020.2986024.

[2] M. Khan, F. G. Glavin, and M. Nickles, "Federated Learning as a Privacy Solution - An Overview," Procedia Comput. Sci., vol. 217, pp. 316–325, Jan. 2023, doi: 10.1016/J.PROCS.2022.12.227.

[3] D. C. Nguyen, M. Ding, P. N. Pathirana, A. Seneviratne, J. Li, and H. Vincent Poor, "Federated Learning for Internet of Things: A Comprehensive Survey," IEEE Commun. Surv. Tutorials, vol. 23, no. 3, pp. 1622–1658, Jul. 2021, doi: 10.1109/COMST.2021.3075439.

[4] M. Abdel-Basset, H. Hawash, N. Moustafa, I. Razzak, and M. Abd Elfattah, "Privacy-preserved learning from non-iid data in fog-assisted IoT: A federated learning approach," Digit. Commun. Networks, Dec. 2022, doi: 10.1016/J.DCAN.2022.12.013.

[5] X. Gu, Z. Tianqing, J. Li, T. Zhang, W. Ren, and K. K. R. Choo, "Privacy, accuracy, and model fairness trade-offs in federated learning," Comput. Secur., vol. 122, p. 102907, Nov. 2022, doi: 10.1016/J.COSE.2022.102907.

[6] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated Learning: A Survey on Enabling Technologies, Protocols, and Applications," IEEE Access, vol. 8, pp. 140699–140725, 2020, doi: 10.1109/ACCESS.2020.3013541.

[7] Y. Guo, F. Liu, T. Zhou, Z. Cai, and N. Xiao, "Seeing is believing: Towards interactive visual exploration of data privacy in federated learning," Inf. Process. Manag., vol. 60, no. 2, p. 103162, Mar. 2023, doi: 10.1016/J.IPM.2022.103162.

[8] T. Chang, L. Li, M. H. Wu, W. Yu, X. Wang, and C. Z. Xu, "PAGroup: Privacy-aware grouping framework for high-performance federated learning," J. Parallel Distrib. Comput., vol. 175, pp. 37–50, May 2023, doi: 10.1016/J.JPDC.2022.12.011.

[9] S. Banabilah, M. Aloqaily, E. Alsayed, N. Malik, and Y. Jararweh, "Federated learning review: Fundamentals, enabling technologies, and future applications," Inf. Process. Manag., vol. 59, no. 6, p. 103061, Nov. 2022, doi: 10.1016/J.IPM.2022.103061.

[10] W. Huang, T. Li, D. Wang, S. Du, J. Zhang, and T. Huang, "Fairness and accuracy in horizontal federated learning," Inf. Sci. (Ny)., vol. 589, pp. 170–185, Apr. 2022, doi: 10.1016/J.INS.2021.12.102.

[11] S. Feng, "Vertical federated learning-based feature selection with non-overlapping sample utilization," Expert Syst. Appl., vol. 208, p. 118097, Dec. 2022, doi: 10.1016/J.ESWA.2022.118097.

[12] Z. Yao and C. Zhao, "FedTMI: Knowledge aided federated transfer learning for industrial missing data imputation," J. Process Control, vol. 117, pp. 206–215, Sep. 2022, doi: 10.1016/J.JPROCONT.2022.08.004.

[13] J. Chen, J. Xue, Y. Wang, L. Huang, T. Baker, and Z. Zhou, "Privacy-Preserving and Traceable Federated Learning for data sharing in industrial IoT applications," Expert Syst. Appl., vol. 213, p. 119036, Mar. 2023, doi: 10.1016/J.ESWA.2022.119036.

[14] L. Li, Y. Fan, M. Tse, and K. Y. Lin, "A review of applications in federated learning," Comput. Ind. Eng., vol. 149, p. 106854, Nov. 2020, doi: 10.1016/J.CIE.2020.106854.

[15] B. Li, Y. Wu, J. Song, R. Lu, T. Li, and L. Zhao, "DeepFed: Federated Deep Learning for Intrusion Detection in Industrial Cyber-Physical Systems," IEEE Trans. Ind. Informatics, vol. 17, no. 8, pp. 5615–5624, Aug. 2021, doi: 10.1109/TII.2020.3023430.

[16] S. Agrawal et al., "Federated Learning for intrusion detection system: Concepts, challenges and future directions," Comput. Commun., vol. 195, pp. 346–361, Nov. 2022, doi: 10.1016/J.COMCOM.2022.09.012.

[17] W. Zhang et al., "Semi-asynchronous personalized federated learning for short-term photovoltaic power forecasting," Digit. Commun. Networks, Mar. 2022, doi: 10.1016/J.DCAN.2022.03.022.

[18] I. Ullah, U. U. Hassan, and M. I. Ali, "Multi-level Federated Learning for Industry 4.0 - A Crowdsourcing Approach," Procedia Comput. Sci., vol. 217, pp. 423–435, Jan. 2023, doi: 10.1016/J.PROCS.2022.12.238.

[19] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A Joint Learning and Communications Framework for Federated Learning over Wireless Networks," IEEE Trans. Wirel. Commun., vol. 20, no. 1, pp. 269–283, Jan. 2021, doi: 10.1109/TWC.2020.3024629.

[20] W. Oh and G. N. Nadkarni, "Federated Learning in Health care Using Structured Medical Data," Adv. Kidney Dis. Heal., vol. 30, no. 1, pp. 4–16, Jan. 2023, doi: 10.1053/J.AKDH.2022.11.007.

[21] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and Privacy-Enhanced Federated Learning for Industrial Artificial Intelligence," IEEE Trans. Ind. Informatics, vol. 16, no. 10, pp. 6532–6542, Oct. 2020, doi: 10.1109/TII.2019.2945367.

[22] A. Yang et al., "Review on application progress of federated learning model and security hazard protection," Digit. Commun. Networks, Nov. 2022, doi: 10.1016/J.DCAN.2022.11.006.

[23] Y. Zhao, J. Zhang, and Y. Cao, "Manipulating vulnerability: Poisoning attacks and countermeasures in federated cloud–edge–client learning for image classification," Knowledge-Based Syst., vol. 259, p. 110072, Jan. 2023, doi: 10.1016/J.KNOSYS.2022.110072.

[24] Z. Chen, P. Tian, W. Liao, and W. Yu, "Towards multi-party targeted model poisoning attacks against federated learning systems," High-Confidence Comput., vol. 1, no. 1, p. 100002, Jun. 2021, doi: 10.1016/J.HCC.2021.100002.

[25] Z. Zhang, Y. Zhang, D. Guo, L. Yao, and Z. Li, "SecFedNIDS: Robust defense for poisoning attack against federated learning-based network intrusion detection system," Futur. Gener. Comput. Syst., vol. 134, pp. 154–169, Sep. 2022, doi: 10.1016/J.FUTURE.2022.04.010.

[26] N. Rodríguez-Barroso, D. Jiménez-López, M. V. Luzón, F. Herrera, and E. Martínez-Cámara, "Survey on federated learning threats: Concepts, taxonomy on attacks and defences, experimental study and challenges," Inf. Fusion, vol. 90, pp. 148–173, Feb. 2023, doi: 10.1016/J.INFFUS.2022.09.011.

[27] A. K. Nair, E. D. Raj, and J. Sahoo, "A robust analysis of adversarial attacks on federated learning environments," Comput. Stand. Interfaces, vol. 86, p. 103723, Aug. 2023, doi: 10.1016/J.CSI.2023.103723.

[28] Y. Gu, Y. Bai, and S. Xu, "CS-MIA: Membership inference attack based on prediction confidence series in federated learning," J. Inf. Secur. Appl., vol. 67, p. 103201, Jun. 2022, doi: 10.1016/J.JISA.2022.103201.

[29] K. Wei et al., "Federated Learning with Differential Privacy: Algorithms and Performance Analysis," IEEE Trans. Inf. Forensics Secur., vol. 15, pp. 3454–3469, 2020, doi: 10.1109/TIFS.2020.2988575.