JCST Journal of Data Acquisition and Processing

ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.776863

FEDERATED LEARNING APPLICATIONS, CHALLENGES & FUTURE DIRECTIONS – A REVIEW

A. Shubha[1], Dr A. Kanagaraj[2], S. Sathiyapriya[3], N. Balakumar[4], P. Karthiga[5]
[1,3,4,5] Ph.D Research Scholar, Department of Computer Science, Nallamuthu Gounder Mahalingam College
Pollachi, Tamil Nadu - 642001, India. Email: as.shuba68@gmail.com
[2] Assistant Professor, Department of Computer Science, Nallamuthu Gounder Mahalingam College
Pollachi, Tamil Nadu - 642001, India. Email: a.kanagaraj@gmail.com

ABSTRACT

Federated learning (FL) is a new technology that has been a hot research topic. It enables the training of an algorithm across multiple decentralized edge devices or servers holding local data samples without exchanging them. Federated Learning embodies the principles of focused minimization and data collection, costs and can reducing privacy risks, and centralized machine learning approaches. Motivated by the growth in Federated Learning research, this thesis considers advances and presents many collections of challenges and open problems. This monograph describes the defining characteristics and challenges of the Federated Learning setting, highlights important practical considerations, constraints, and then enumerates a range of valuable research directions. The goals of this work are to highlight research problems that are of significant theoretical and practical interest, and to encourage research on problems that could have significant real-world impact. Finally, the paper highlights the limitations present in recent works and presents some future directions for this technology.

Keywords: *Federated Learning, Decentralized Machine Learning, Smart Health Care.*

1. INTRODUCTION

Since the concept of AI was put forward in 1956, AI technology has more and more profound impact on human life [1,2]. As great progress has been made in AI technology in recent years, various application fields have stepped into intelligence [3, 4]. On the road of AI development, models, computing power, chip performance, and other technical issues have been the focus of academic research, so that AI technology can continue to evolve. For machines to truly approach the level of human thought, they need to be trained with vast amounts of real data [5, 6]. However data security, cloud computing power, data silos and other risks will inevitably become constraints for Artificial Intelligence to win collect private data, user trust and achieve largescale implementation [7].Federated Learning was used in most scenarios, researchers have found that there are still many challenges to be addressed. For example, a large amount of research work has realized that federated learning, originally intended to protect privacy, is more vulnerable to attacks by malicious nodes in many practical scenarios than traditional deep learning frameworks [76]. Federated Learning was first introduced google in 2016 [8]. According to Google, FL was first used on the google keyboard, mainly to protect users' private data [9] and to improve the language model quality [10]. In the year of 2017 Google propose FL is a product-development area, formal research and in the year of March 2019 Google Brain told the availability of TensorFlow Federated (TFF) as a standard tool-set for Federated Learning. As only the parameters, which do not reveal the client's identity, are collected by a federated learning server, the anonymous clients may contain attackers that upload malicious data to the server without exchange the data itself [77-80]. Originally Created for various devices, such as edge device and mobile [11] Data privacy in a decentralized collaborative learning scenario is an important consideration in this area of research, which is why it has been linked to in [12] Users are highly concerned about their data privacy, and therefore, acquiring and using personal data is very challenging [13]. privacy has been a major concern in the adoption of AI techniques in commercial environments. Recently, most digital applications such as contact tracing apps developed for containing the spread of the novel coronavirus disease 2019 (COVID-19) were not welcomed by many people across the globe due to privacy concerns [14]. has is realizing a answer for cost-effective intelligent IIoT applications with improved privacy protection. Motivated by these appealing characteristics, a flurry of research actions joining FL with IIoT [14–16]. However, these works only focus on certain application domains in IIoT, such as cognitive computing [14], industrial artificial intelligence [15], and digital twin-enabled IIoT [16], while a holistic overview on the use of FL in key IIoT services and applications is still missing. we discuss the roles of FL in key industrial IIoT applications, including transportation, smart manufacturing, smart grid, and healthcare. We also highlight interesting open issues for future FL IIoT research. Now, FL tool is still in continuous betterment. By joining the existing literature, now FL mainly faces three problems: security threats and privacy, enormous communication and variety challenges overhead of FL [14]. we analyzed the current research status and prospect for the development of FL in the future.

The involvements of this paper to the FL research are as follows:

1. It provides an in-depth survey and analysis of the latest papers on FL.
2. It Worries the practical application prospects of FL and the challenges faced in the process of practical applications, and has unique views on the current development and future prospects of FL.

2. FEDERATED LEARNING

Federated learning is a new technique that trains an algorithm across multiple decentralized edge devices or servers holding local data samples, without exchanging them. This approach stands in contrast to traditional centralized machine learning techniques where all the local datasets are uploaded to one server, as well as to more classical decentralized approaches which often assume that local data samples are identically distributed. Federated learning is used for distributed training of machine learning algorithms on multiple edge devices without exchanging training data. Therefore, Federated learning introduces a new learning paradigm where statistical methods are trained at the edge in distributed networks. decentralized data and at the same time protect user confidential data by design. [17]

FL methods are usually subdivided into Horizontal Federated Learning (HFL), Vertical Federated Learning (VFL) and Federated Transfer Learning (FTL), which are suitable for solving different practical problems. Shown Fig. 1. Then, all category is briefly explained.

2.1 Horizontal Federated Learning

Horizontal federated learning, or sample-based federated learning, is introduced in the scenarios in which datasets share the identical feature space but different space in samples (Figure 1(a)). For example, two regional banks may have very different user groups from their respective regions, and the intersection set of their users is very small. However, their business is very similar, so the feature spaces are the same. The authors of [18] proposed a collaboratively deep-learning scheme in which player develop independently and distribute only subsets of updates of parameters. In the year of 2017, Google introduce a Horizontal Federated-Learning solution to Android phone model updates [19]. federated learning attracts great attention from plenty of applications. Like wake word detection [20], emoji prediction, personalized model training, Internet of Things [81–85]. For the dataset of each client, if the overlap of data features is larger than that of users, that is, there are more of the same data features and fewer of the same users, this is called horizontal federated learning (HFL) [21]. HFL is leads by the feature dimension of the data, and takes out the parts with the same characteristics of the participants but different users for joint training. In the process of HFL, the training sample space is enlarged by the sample union between participants, thereby improving the accuracy and generalization ability of model [22]. According to [23], several sites may execute independent tasks while maintaining security and sharing knowledge by using a multitask-style FL system. High fault toleration strayer, communication costs difficulties may be all talked by their multitask learning. It was suggested in [24] to establish a safe client–server structure where data is partitioned by users as well as models developed on client devices work together to produce a global federated algorithm in the framework of an interconnected federated learning system. The model-building procedure prevents any data from leaking

We summarize horizontal federated learning as

$$Xi = Xj,\ Yi = Yj,\ Ii \neq Ij,\ \forall Di,\ Dj,\ i \neq j. \qquad (1)$$

Security Definition. It is typically assuming honest participants and security against an honest-but-curious server [25,26]. That is, only the server can compromise the privacy of data participants. Security proof has been provided in these works. Recently, another security pattern views vicious users [27] was also proposed, additional privacy challenges. At the end of the practice, the universal model and all of the model parameters are exposed to all participants.

2.2 Vertical Federated Learning

VFL are largely different between user. For dataset of user, the overlap of users is greater than that of data features, that is, each client dataset has more of the same users, and the data features are hardly duplicated [28]. It is show in (Figure 1(b)) is applicable to the cases in which two datasets share the same sample ID space but differ in feature space. For example, consider two different companies in the same city: one is a bank and the other is an ecommerce company. Their user sets are likely to contain most of the residents of the area; thus, the intersection of their user space is large. However, since the bank records the user's revenue and expenditure behavior and credit rating and the e-commerce retains the user's browsing and purchasing history, their feature spaces are very different. Suppose that we want both parties to have a prediction model for product purchases based on user and product information. Under such a federal mechanism, the identity and the status of each participating party is the same, and the federal system helps everyone establish a "common wealth" strategy, which is why this system is called federated learning

$$Xi \neq Xj,\ Yi \neq Yj,\ Ii = Ij\ \forall Di,\ Dj\ ,i \neq j. \qquad (2)$$

Security Definition. A vertical federated-learning assumes honest but curious contributor. E.g. the two group are non-colluding and at most one is compromised by an adversary. The security definition is that the adversary can learn data only from the client that it corrupted and not data from the other client beyond what is revealed by the input and output. SMC provides formal privacy proof for these protocols [29]. finally, each party holds only those model parameters associated to its own features. Therefore, at inference time, the two parties also need to collaborate to generate output

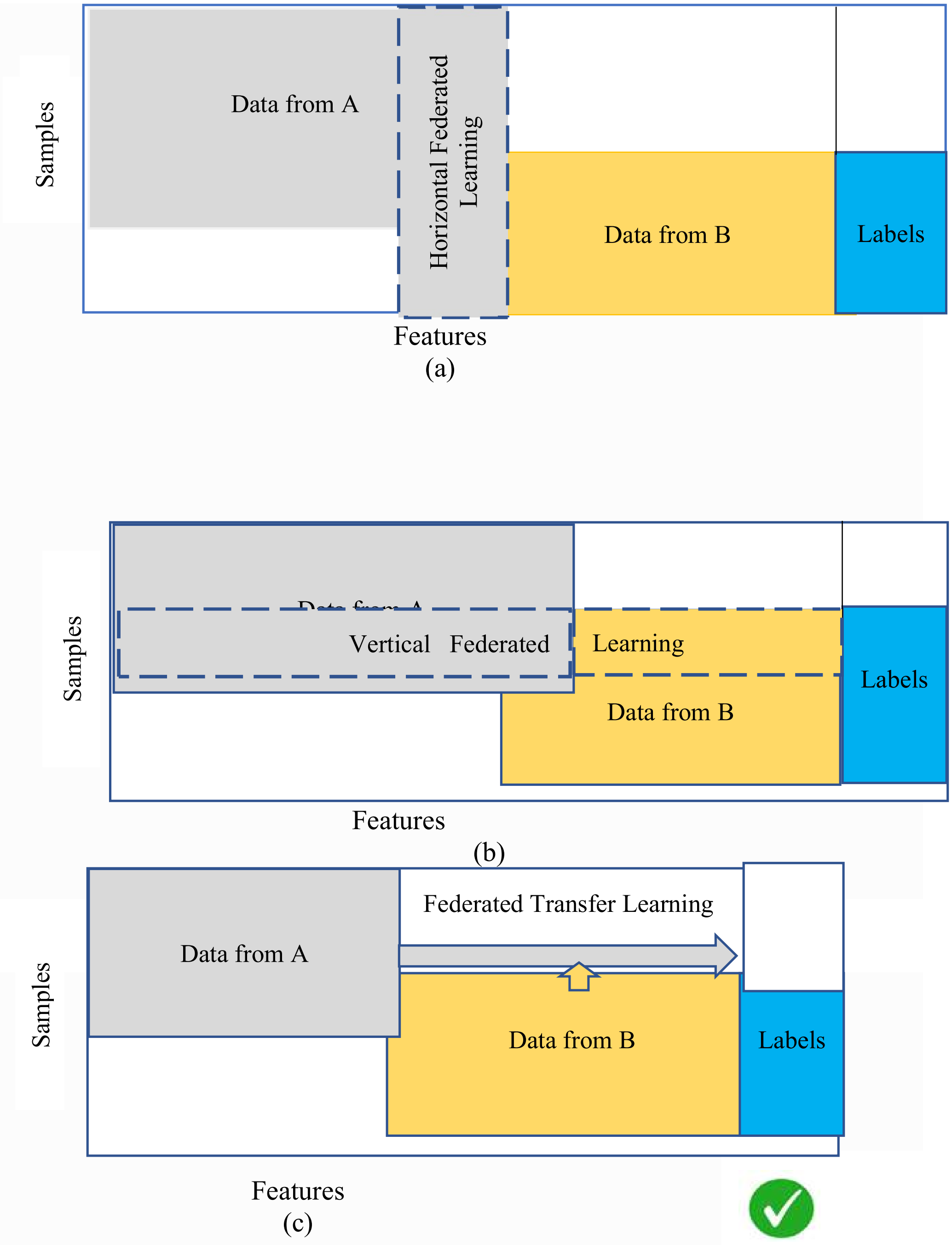

Fig 1. (a) Horizontal Federated Learning (b) Vertical Federated Learning (c) Federated Transfer Learning

2.3 Federated Transfer Learning

Federated transfer learning applies to the scenarios in which two datasets differ not only in samples but also in feature space. Consider two institutions: one is a bank located in China and the other is an e-commerce company located in the United States. Owing to geographical restrictions, the user groups of the two institutions have a small intersection. rather, owing to the different businesses, only a small portion of the feature space from both parties overlaps. In this case, transfer-learning [30] techniques can be applied to provide solutions for the entire sample and feature space under a federation (Figure 1(c)). Specially, a common representation between the two feature spaces is learned using limited common sample sets and later applied to obtain predictions for samples with only one-side features. FL algorithms:

$$Xi \neq Xj,\ Yi \neq Yj,\ Ii \neq Ij\ \forall Di,\ Dj,\ i \neq j. \tag{3}$$

Security Definition. A federated transfer learning system typically involves two parties. As will be discusses in the next part, its protocols are same to the ones in this learning, in which case the security definition for vertical federated learning can be extended here.

3. FEDERATED LEARNING APPLICATION

In the big data environment, users' personal information and behavioral data can be recorded in mobile smart devices or edge servers. So, data privacy protection provides a necessary guarantee for intelligent development, and various fields pay more attention to the security and privacy of private information. FL has acted valuable place in the fields of various Fields. FL application areas are smart health care, banking and finance, smart industry, smart city and urban services. Now we explain in detail the applications of FL in these classic scenarios.

3.1 FL for Smart Healthcare

In the medical field, individuals usually pay more attention to the privacy information of patients can not be disclosed, at the same time, various hospitals cannot exchange patient information privately, so it is quite suitable to use FL technique for protecting pathological information [31].Data from Electronic Health Records (EHRs) has become an invaluable resource for biological research [32], Electronic health records, scientific journals, and similar archive on the internet, visual, and audio consultation with doctor, or internet-based process to connect with medical personal, give feedback to doctors, transfer test results, and so on are examples of e-health systems that can revolutionize the healthcare services they provide. [33]. To build stronger Al models for medical tasks, such as medical imaging [34]. For example, the classification tasks of COVID19 detections [35], cancer diagnoses [36] and autism spectrum disorder (ASD) [37] are considered in a FL setting in healthcare. A privacy-preserving framework for learning about patient similarity in through various institutions was described by Lee et al. [38]. To locate comparable patients in other hospitals, their approach does not require sharing any patient-level data. In a federated learning the authors of [39] employed models for electronic health records

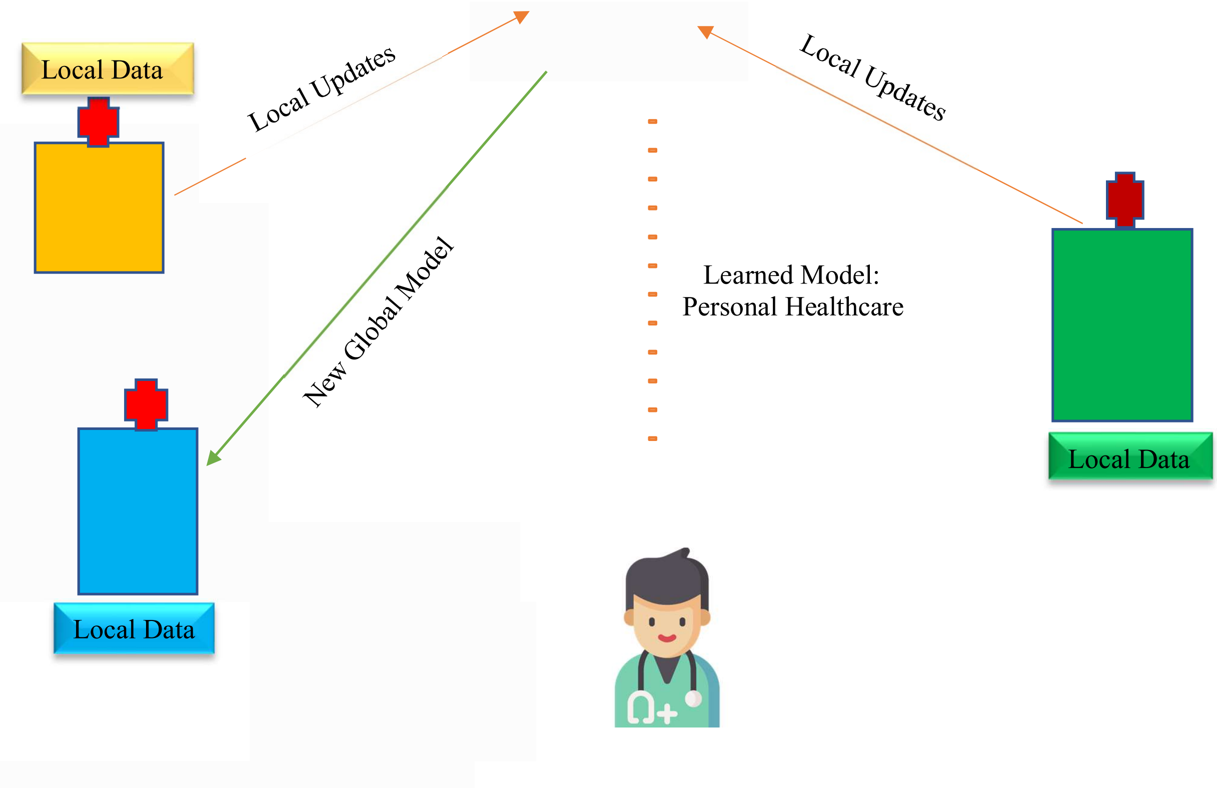

Fig 2. FL in Healthcare

3.2 FL for Open Banking and Finance

Open banking is an emerging trend in turning banks into financial service platforms, namely banking as a service. From a financial technology perspective, open banking refers to:[40] 1) use of Application Programming Interfaces (APIs) this enables third party developers to build applications and services around the financial institution, 2) greater financial transparency options for account holders ranging from open data to private data, and 3) the use of open-source technology to achieve the above.[41]. It has significantly advanced along various pathways [42]. However, there are inherent risks in sharing banking data, which is sensitive, privacy-concerned, and valuable. It is critical to developing processes and governance underpinning the technical connections [40]. To solve practical challenges, such as limited times to access personal data, broad heterogeneity across users, narrow scope of one user, and managing incentives for data contributors. Open innovation is “a distributed innovation using pecuniary, process based on purposively managed knowledge flows across organizational boundaries, and nonpecuniary mechanisms in line with the organization’s business model” [43]. The flows of information may involve various ways to leverage internal and external resources and knowledge. This is a kind of open innovation in the banking industry. By leveraging both external and internal knowledge, many innovative applications will emerge to benefit the whole finance industry, including both banks and third-party companies

3.3 FL for Smart Manufacturing

This is a response to the new sensors, big data and communication technologies. The core concept of SM can be described as the creation of data-driven, intelligent, and flexible manufacturing operations by fully integrating them at all levels and stages of manufacturing, using smart technologies and applying cyber-physical systems and IIoT. [44,45] SM is capable of operating in a dynamic and changing environment, providing an enhanced product quality by meeting the products design while being responsive to customers' demands and improving its sustainability, for instance, by reducing the wastes and increasing the accuracy of the manufacturing processes.[46].Manufacturing of smart vehicles in automotive industry must be capable of employing IIoT along with ML techniques paving the way for smart industry experience enabling efficient and sustainable production [47] in various applications are development of many fields including governance, education, telecommunication, healthcare, finance, smart manufacturing, etc. Taking the example of smart manufacturing, FL could be applied to improve the complete journey of a product from initial design, development, service, and disposal. In the smart automotive industry, hundreds (or sometimes thousands) of workstations, machines, and sensors are connected and require the deployment of ML models to effectively perform the assigned tasks such as robotic arms, fault detection, or object recognition [48]. FL can bring a lot of benefits to the entire product life cycle of the automotive industry for large companies as well as small enterprises by giving them access to full intelligence.

Smart manufacturing refers to the integration of intelligence into manufacturing processes where AI techniques play important roles in learning big data generated from industrial machines for process modeling, monitoring, prediction and control in production stages. The AI functions often require data sharing among manufacturers and factories which is not an ideal solution due to growing user privacy concerns. FL can understand intelligence for industrial systems without data exchange, by the collaborative data learning of distributed industrial devices and machines. Given the fact that there are diverse industrial services in industries, e.g., production monitoring with robots, product assembly with automatic manipulator arms, package delivery and logistics with vehicles, it is desired to develop a multiple FL services solution to deal with different industrial services in the co-working IIoT ecosystem.[49]

3.4 FL for Agricultural

As the global population continues to grow, the amount of food consumed worldwide is increasing as well. It is not an easy task to rise food yield capacity as demand increases, and the challenge is compounded when we have to also consider the diverse requirements of different multicultural societies. Hence, there have been attempts to find ways of optimizing food production, for example by maximizing the utility of water and land [50]. In other words, sustainability in agriculture can be ensured through the appropriate use of data in decision-making. [51]. This is a complex and highly valuable sector in the world economy, yet the hostility that arises from competitive advantage has snuffed the possibility of collaboration and openness in data sharing that has the potential to benefit all parties [52,53]. Data sharing can help address the unsafe food vertically through the supply chain. Substantial work has been done to address the traceability of food and drink with added pressure from consumer demands [54].

3. 5 FL for Smart Grid.

Smart grid plays an integral part in building smart city environment which not only provides energy resources to smart city applications such as transportation, manufacturing, but also has impacts on environmental, security, and social aspects in Industry 4.0. FL can enable intelligent solutions for smart grid management and energy transmissions in a decentralized manner while helping promote privacy. [55]. When FL is applied to smart city construction, government agencies and private enterprises will act as data owners, that is, clients. And the urban data management center will become the service. To solve various problems in the construction of smart cities, the service will unite with different levels of institutions in the city to train a global model, it will deal with different tasks, so as to provide more convenient urban services for citizens [56]

4. CHALLENGES IN FEDERATED LEARNING

4.1 **Security and Privacy**

Privacy and security are a top concern of federated learning. According to [57], most research of model attacks will assume that the attacker can barely access the model input, because the data for the training model remains internally confidential. Although security and privacy are often considered equivalent intuitively, the difference actually exists. As mentioned above, privacy protection generally refers to the non-public exposure of sensitive personal information. Such information may be the user's health information, travel trajectories, salary level, etc.

4.2 Communication Challenge

In a real world scenario, there can be millions of devices involved in the network, and each device may spend far less time training model locally than the network communication[58].Federated learning's communication costs should be taken into account as well, since network speeds cannot be guaranteed. As a result, the development of systems with great communication efficiency is required in order to create FL practicable. Communication at a high-cost data created on each device must stay local due to communication being a significant bottleneck in federated networks.[59]

4.3 Training-Data-Related Challenges

This are the most important of FL. There exist multiple challenges with regard to the quality of data. In addition, the privacy of the training data is one of the hot challenges in the FL [60]. Recently, the non-i.i.d. nature of the training data poses various technical challenges in the FL paradigm, and their solution has become more urgent than ever. In addition, guaranteeing the quality of data and preventing it from poisoning the paradigm is also one of the main challenges. To exactly benefit from the potential of FL, training-data-related challenges need robust solutions

4.4 Client-Related Challenges

In the FL, clients are regarded as independent, which means they can perform most activities autonomously. Hence, they can leave the system at any time, which can lead to longer convergence and disturbs the training process [61]. The prevention of the client's dropout is a longstanding challenge in FL

4.5 Servers-Related Challenges

The server is answerable for the orchestration of the aggregating models, local models, and sharing the global model. However, in some cases, multiple attacks can be executed on the server by adversaries, which makes the FL system untrustworthy. Since the server is only concerned with the model weights without deep inspection, it cannot filter malicious clients, which degrades the performance of the FL paradigm

4.6 Systems Heterogeneity

The computational, storage, and communication capabilities of every device in federated networks may differ due to variability in hardware, network connectivity (3G, 4G, 5G, WIFI), and power (battery level). Additionally, the network size and systems-related constraints on each device typically result in only a small fraction of the devices being active at once. Heterogeneity in systems Due to the variety of people's devices, network status, storage, and processing capability of devices, the training process of computing and communication capabilities will be different[62].

4.7 Unreliable Model Upload

Nodes might intentionally or accidentally mislead the server [63] while aggregating global models in FL. Using malicious model parameters, an attacker may cause model training to go awry by altering the aggregate of the global model. Mobile devices might also upload low-quality models because of the unpredictable network environment, which would have a negative impact on FL. For FL, it is essential to avoid uploading an unstable local model is essential for applying supervised learning mechanisms in a dataset.

4.8 Intrusion Detection

An intrusion detection is a device that monitors outbound and inbound network traffic, continuously analyzing activity for changes in patterns, and alerts an administrator when it detects unusual behavior. An administrator then evaluation alarms and bring actions to remove the threat. As the number of cyberattacks and intrusions continue to rise, monitoring and securing your company's network has never been more pressing. In 2021 alone, the FBI's Internet Crime Complaint Center (IC3) received more than 800,000 complaints about data breaches, malware and more. These complaints totaled nearly $7 billion in losses—and they only represent the reported cases. If you want to protect yourself and your business from these threats, you need a comprehensive cybersecurity setup. There are several kinds of IDS solutions on the market today.

5. FUTURE DIRECTION

The study of federated learning is one that is both current and continuous. This section's survey of related and current work has started to address some of the issues raised, but there are still a number of important questions that need to be answered. Our prior obstacles (privacy and security, Extreme communication, statistical heterogeneity, heterogeneity and concerns) are briefly reviewed in this section, as are new challenges such as productionizing and benchmarking in federated contexts, which are described in the next paragraphs.

5.1 Privacy and Security

As mentioned above, although federated learning was widely used, it still faces many challenges. Among issues, we believe that learning how to better protect users' privacy will be one of the most important points in the future development of federated learning. With the establishment of the EU General Data Protection Regulations (GDPR) in 2018 [64], Recently, in the COVID-19 outbreak, some researchers have suggested that federated learning could better assist in the patient diagnosis while protecting the privacy of medical data [65,66]. Future research should focus on developing strategies for dealing with mixed (sample-specific or device-specific) privacy limitations.

5.2 Extreme Communication Schemes

The extent to which federated learning will require communication is yet to be determined. Machine learning optimization approaches may accept a lack of accuracy; in fact, this imperfection can aid generalization [67]. There has been some research on classic data center communication systems like one-shot or divide-and-conquer [68], but the behavior of these approaches in huge and highly heterogeneous networks is not well known.

5.3 Heterogeneity Diagnostics

Recently, metrics such as local unsimilarity have been used to attempt to value statistical heterogeneity (as developed in the environment of FL and utilized elsewhere in works like [69]). Prior to training, it was impossible to readily compute these metrics across the federated network. The following questions are prompted by the relevance of these metrics: The amount of heterogeneity in federated networks is difficult to predict before they are deployed. Is there a way to measure the level of system-related heterogeneity? It is possible to use existing or new definitions of heterogeneity, both experimentally and conceptually, to develop new federated optimization algorithms with enhanced convergence

5.4 Beyond Supervised Learning

As a reminder, all of the approaches presented so far were designed for supervised learning, which means that they presume that each and every data in the FL network has already been labeled. Realistic federated networks may create data that is either unlabeled or poorly tagged. More sophisticated tasks, such as reinforcement learning, may include exploratory data analysis and aggregate statistics rather than just fitting a model to the available data as described in [70]. Problems in federated networks beyond supervised learning will most likely need to tackle comparable issues of heterogeneity, scalability, and privacy in order to be effectively addressed.

5.5 Aggregation Techniques

Developers who wish to implement FL solutions can benefit from toolkits that offer standardized and preconfigured aggregation algorithms that are suitable for their specific application areas and use cases. Similar to Auto ML solutions, such a toolkit for FL can lower the barrier of entry for no specialist developers.

5.6 Novel models of Asynchrony

Two communication projects commonly studied in distributed optimization are bulk synchronous and asynchronous approaches. However, in federated networks, every device is often undedicated to the task at hand and most devices are not active on any given iteration. Can we devise device-centric communication models beyond synchronous and asynchronous training, where each device can decide when to interact with the server [71].

5.7 Productionizing Federated Learning

FL is put into production. Concerns including idea drift, diurnal fluctuations [72], as well as cold-start concerns (the network expands as additional devices are added) must be addressed with caution. There are a number of actual system-related challenges in production federated learning systems, which we direct readers to in [73].

5.8 Benchmarks

Because federated learning is still in starting stage, we have an opportunity to define its future by ensuring that it is based on real-world circumstances, assumptions, and data sets. Finally, research groups need to improve on current implementations and benchmarking tools, such as

TensorFlow Federated [74] and LEAF [75], in order to make empirical findings more reproducible and to disseminate new federated learning solutions.

6. CONCLUSION

FL technique, as one of the important solutions of privacy protection, has gained rapid development and enough attention in recent years. From the basic knowledge of FL for solving above challenges, researchers have carried out careful research on the various branches of FL, and have done a lot of research work in each branch. Researchers have a realize of FL, and based a perfect combination of FL with various applications.

In this paper Explained the current research achievements of FL, systematically present the concept of FL, the challenges faced by the development of FL, and the research direction of FL combined with various applications. On the basis of these, we made a deep thinking and analysis on the development of Federated Learning in the future and the bottleneck problems to be broken. Federated Learning, future research will continue to focus on security and privacy protection mechanism, client cooperation training mode and robustness, fairness, personalized federated learning mechanism, so as to facilitate the deployment and application of FL technology for in-depth exploration.

References

1. Zhang Z, Zhao M et al (2022) An efficient interval many-objective evolutionary algorithm for cloud task scheduling problem under uncertainty. Inf Sci 583:56–72
2. Wang H, Xie F, Li J, Miu F (2022) Modelling, simulation and optimisation of medical enterprise warehousing process based on FlexSim model and greedy algorithm. Int J Bio-Inspired Comput 19(1):59–66
3. Ren Y, Sun Y et al (2019) Adaptive Makeup Transfer via Bat Algorithm. Mathematics 7(3):273
4. Yang Y, Cai J, Yang H, Zhao X (2021) Density clustering with divergence distance and automatic center selection. Inf Sci 596:414–438
5. Hemalatha B, Rajkumar N (2021) A modified machine learning classification for dental age assessment with effectual ACM-JO based segmentation. Int J Bio-Inspired Comput 17(2):95–104
6. Cui Z, Zhao P et al (2021) An improved matrix factorization-based model for many-objective optimization recommendation. Inf Sci 579:1–14
7. Kuze N, Ishikura S et al (2021) Classification of diversified web crawler accesses inspired by biological adaptation. Int J Bio-Inspired Comput 17(3):165–173
8. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y. Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, ser. Proceedings of Machine Learning Research, A. Singh and J. Zhu, Eds., vol. 54. PMLR, 20–22 Apr 2017, pp. 1273–1282. [Online]. Available: https://proceedings.mlr.press/v54/mcmahan17a.html
9. Yang T, Andrew G, Eichner H, Sun H, Li W, Kong N,Ramage D, Beaufays F. Applied federated learning: Improving google keyboard query suggestions. 2018, arXiv preprint arXiv: 1812.02903
10. Hard A, Rao K, Mathews R, Ramaswamy S, Beaufays F, Augenstein S,Eichner H, Kiddon C, Ramage D. Federated learning for mobile keyboard prediction. 2018, arXiv preprint arXiv: 1811.03604

11. Kairouz, P. et al. Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977 (2019).
12. V. Smith, C.-K. Chiang, M. Sanjabi and A.S. Talwalkar, federated multi-task learning, Advances in Neural Information Processing Systems 30
13. Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST) 10(2), 1–19 (2019)
14. Chan, E.Y.; Saqib, N.U. Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. *Comput. Hum. Behav.* 2021, *119*, 106718.
15. M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and Privacy-Enhanced Federated Learning for Industrial Artificial Intelligence," IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6532–6542, Oct. 2020.
16. W. Sun, S. Lei, L. Wang, Z. Liu, and Y. Zhang, "Adaptive Federated Learning and Digital Twin for Industrial Internet of Things," IEEE Transactions on Industrial Informatics, pp. 1–1, Oct. 2020
17. Y. Qu, S. R. Pokhrel, S. Garg, L. Gao, and Y. Xiang, "A Blockchained Federated Learning Framework for Cognitive Computing in Industry 4.0 Networks," IEEE Transactions on Industrial Informatics, pp. 1–1, Jul. 2020.
18. Reza Shokri and Vitaly Shmatikov. 2015. Privacy-preserving deep learning. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS'15). ACM, New York, NY, 1310–1321. DOI:https: //doi. org/10.1145/2810103.2813687
19. Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2017. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS'17). ACM, New York, NY, 1175–1191
20. Leroy D, Coucke A, Lavril T, Gisselbrecht T, Dureau J. Federated learning for keyword spotting. In: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. 2019, 6341− 6345
21. Zhang K, Song X, Zhang C, Yu C (2022) Challenges and future directions of secure federated learning: a survey. Front Compute Sci 16(5):165817
22. Feng C, Liu B et al (2022) Blockchain-Empowered Decentralized Horizontal Federated Learning for 5G-Enabled UAVs. IEEE Trans Industry Inf 18(5):3582–3592
23. V. Smith, C.-K. Chiang, M. Sanjabi, and A. S. Talwalkar,Federated multi-task learning, Advances in neural information processing systems, vol. 30, 2017
24. H. B. McMahan, E. Moore, D. Ramage, and B. A. y Arcas, Federated learning of deep networks using model averaging, arXiv preprint arXiv:1602.05629, 2016.
25. Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. 2017. Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS'17). ACM, New York, NY, 1175–1191. DOI:https://doi.org/10.1145/3133956.3133982
26. Le Trieu Phong, Yoshinori Aono, Takuya Hayashi, Lihua Wang, and Shiho Moriai. 2018. Privacy-preserving deep learning via additively homomorphic encryption. IEEE Trans. Information Forensics and Security 13, 5 (2018), 1333– 1345.

27. Briland Hitaj, Giuseppe Ateniese, and Fernando Pérez-Cruz. 2017. Deep models under the GAN: Information leakage from collaborative deep learning. CoRR abs/1702.07464 (2017).
28. Dai M, Xu A, Huang Q, Zhang Z, Lin X (2021) Vertical federated DNN training. Phys Communication 49:101465
29. O. Goldreich, S. Micali, and A. Wigderson. 1987. How to play any mental game. In Proceedings of the 19th Annual ACM Symposium on Theory of Computing (STOC'87). ACM, New York, NY, 218–229. DOI:https://doi.org/10.1145/ 28395.28420
30. Sinno Jialin Pan and Qiang Yang. 2010. A survey on transfer learning. IEEE Trans. Knowl. Data Eng. 22, 10 (Oct. 2010), 1345–1359. DOI:https://doi.org/10.1109/TKDE.2009.191
31. Xu X, Peng H, Bhuiyan M et al (2022) Privacy-Preserving Federated Depression Detection From Multisource Mobile Health Data. IEEE Trans Industr Inf 18(7):4788–4797
32. B.S. Glicksberg, K.W. Johnson and J.T.Dudley, The next generation of precision medicine: Observational studies, electronic health records, biobanks and continuous monitoring, Human Molecular Genetics 27(R1) (2018), R56–R62.
33. Muhammad, G., Rahman, S.K.M.M., Alelaiwi, A., Alamri, A.: Smart health solution integrating iot and cloud: a case study of voice pathology monitoring. IEEE Commun. Magazine 55(1), 69–73
34. Dinh C. Nguyen, Ming Ding, Pubudu N. Pathirana, Aruna Seneviratne, Jun Li, Dusit Niyato, Fellow, IEEE, and H. Vincent Poor, Fellow, IEEE
35. S. Bharati and M.R.H. Mondal, 12 Applications and challenges of AI-driven IoHT for combating pandemics: a review Computational Intelligence for Managing Pandemics, in: A. Khamparia, M.R.H. Mondal, P. Podder, B. Bhushan, V.H.C.d. Albuquerque and S. Kumar, Eds.: De Gruyter, 2021, pp. 213–230.
36. P. Podder, S. Bharati, M.A. Rahman and U. Kose, Transfer Learning for Classification of Brain Tumor, in: Deep Learning for Biomedical Applications, CRC Press, 2021, pp. 315– 328.
37. M. Sadiq Iqbal, N. Akhtar, A.H.M. Shahariar Parvez, S. Bharati and P. Podder, Ensemble learning-based EEG feature vector analysis for brain computer interface, in: Evolutionary Computing and Mobile Sustainable Networks, Springer, 2021, pp. 957–969.
38. J. Lee, J. Sun, F. Wang, S. Wang, C.-H. Jun and X. Jiang, Privacy-preserving patient similarity learning in a federated environment: Development and analysis, JMIR Medical Informatics 6(2) (2018), e7744.
39. Y. Kim, J. Sun, H. Yu and X. Jiang, Federated tensor factorization for computational phenotyping, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2017, pp. 887–895
40. PYMNTS: Open banking targets smb apps, payments data. Tech. rep., PYMNTS.com (2020)
41. Group, O.B.W., et al.: The open banking standard. Tech. rep., working paper,Open Data Institute (2018)
42. Brodsky, L., Oakes, L.: Data sharing and open banking. McKinsey Company (2017)
43. Chesbrough, H., Vanhaverbeke, W., West, J.: New frontiers in open innovation. Oup Oxford (2014)
44. V. W. C. Fung and K. C. Yung, "An Intelligent Approach for Improving Printed Circuit Board Assembly Process Performance in Smart Manufacturing," International Journal of Engineering Business Management 12 (January 2020), https://doi.org/10.1177/1847979020946189

45. B. Wang, F. Tao, X. Fang, C. Liu, Y. Liu, and T. Freiheit, "Smart Manufacturing and Intelligent Manufacturing: A Comparative Review," Engineering. Published ahead of print, 20 September, 2020, https://doi.org/10.1016/j.eng.2020. 07.017

46. Y. Lu, X. Xu, and L. Wang, "Smart Manufacturing Process and System Automation – A Critical Review of the Standards and Envisioned Scenarios," Journal of Manufacturing Systems 56 (July 2020): 312–325, https://doi.org/10.1016/j.jmsy. 2020.06.010

47. M. Zolanvari, M. A. Teixeira, L. Gupta, K. M. Khan, and R. Jain, "Machine learning-based network vulnerability analysis of industrial internet of things," IEEE Internet of Things Journal, vol. 6, no. 4, pp. 6822–6834, 2019.

48. Boopalan, Parimala, Swarna Priya Ramu, Quoc-Viet Pham, Kapal Dev, Praveen Kumar Reddy Maddikunta, Thippa Reddy Gadekallu, and Thien Huynh-The. "Fusion of Federated Learning and Industrial Internet of Things: A survey." Computer Networks (2022)

49. Dinh C. Nguyen, Ming Ding, Pubudu N. Pathirana, Aruna Seneviratne, Jun Li, Dusit Niyato, Fellow, IEEE, and H. Vincent Poor, Fellow, IEEE

50. Internet of things for the future of smart agriculture: a comprehensive survey of emerging technologies IEEE/CAA J. Autom. Sin. (2021)

51. M.A. Ferrag *et al.*Rdtids: rules and decision tree-based intrusion detection system for Internet-of-things networks

52. ODI, Data sharing in the private sector, https://theodi.org/article/ new-survey-finds-just-27-of-british-businesses-are-sharing-data/ (2020).

53. A. Durrant, M. Markovic, D. Matthews, D. May, G. Leontidis, J. Enright, How might technology rise to the challenge of data sharing in agri-food?, Global Food Security 28 100493.

54. H. Feng, X. Wang, Y. Duan, J. Zhang, X. Zhang, Applying blockchain technology to improve agri-food traceability: A review of development methods, benefits and challenges, Journal of Cleaner Production (2020)

55. A. Taik and S. Cherkaoui, "Electrical Load Forecasting Using Edge Computing and Federated Learning," in Proc. IEEE International Conference on Communications (ICC), Dublin, Ireland, Jun. 2020,

56. Jiang C, Li R, Chen J et al (2021) Modelling the green supply chain of hotels based on front-back stage decoupling: perspective of ant colony labour division. Int J Bio-Inspired Comput 18(2):176–188

57. Ha T, Dang T K, Le H, Truong T A. Security and privacy issues in deep learning: a brief review. SN Computer Science, 2020, 1(5): 253

58. Konečný J, McMahan H B, Yu F X, Richtárik P, Suresh A T, Bacon D.Federated learning: strategies for improving communication efficiency. 2016, arXiv preprint arXiv: 1610.05492 WANG Luping,

59. K. Bonawitz et al., Towards federated learning at scale: System design, Proceedings of Machine Learning and Systems 1(2019), 374–388.

60. Hao, M.; Li, H.; Luo, X.; Xu, G.; Yang, H.; Liu, S. Efficient and privacy-enhanced federated learning for industrial artificial intelligence. *IEEE Trans. Ind. Inform.* 2019, *16*, 6532–6542. [Google Scholar] [CrossRef]

61. Tolpegin, V.; Truex, S.; Gursoy, M.E.; Liu, L. Data poisoning attacks against federated learning systems. In Proceedings of the European Symposium on Research in Computer Security, Guildford, UK, 14–18 September 2020; pp. 480–501. [Google Scholar]

62. Li T, Sahu A K,Zaheer M,Sanjabi M,Talwalkar A,Smith V. Federated optimization in heterogeneous networks.2018,arXiv preprint arXiv: 1812.06127

63. Goddard M. The eu general data protection regulation (GDPR): European regulation that has a global impact. International Journal of Market Research, 2017, 59(6): 703–705

64. H. Eichner, T. Koren, B. McMahan, N. Srebro and K. Talwar, Semi-cyclic stochastic gradient descent, in: International Conference on Machine Learning, PMLR, 2019, pp. 1764– 1773.

65. K. Bonawitz et al., Towards federated learning at scale: System design, Proceedings of Machine Learning and Systems 1 (2019), 374–388.

66. Kumar R, Khan A A, Zhang S, Wang W, Abuidris Y, Amin W, Kumar J. Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging. 2020, arXiv preprint arXiv: 2007.06537

67. Liu B, Yan B, Zhou Y, Yang Y, Zhang Y. Experiments of federated learning for covid-19 chest x-ray images. 2020, arXiv preprint arXiv: 2007.05592

68. Y. Yao, L. Rosasco and A. Caponnetto, On early stopping in gradient descent learning, Constructive Approximation 26(2) (2007), 289–315

69. L. Mackey, M. Jordan and A. Talwalkar, Divide-and-conquer matrix factorization, Advances in Neural Information Processing Systems 24 (2011).

70. D. Yin, A. Pananjady, M. Lam, D. Papailiopoulos, K. Ramchandran and P. Bartlett, Gradient diversity: a key ingredient for scalable distributed learning, in: International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 1998–2007

71. N. Agarwal, A.T. Suresh, F.X.X. Yu, S. Kumar and B. McMahan, cpSGD: Communication-efficient and differentiallyprivate distributed SGD, Advances in Neural Information Processing Systems 31 (2018).

72. Tian Li CSD CMU PUBLISHED November 12, 2019

73. H. Eichner, T. Koren, B. McMahan, N. Srebro, and K. Talwar, Semi-cyclic stochastic gradient descent, in International Conference on Machine Learning, 2019, pp. 1764-1773: PMLR.

74. K. Bonawitz, H. Eichner, and W. Grieskamp, TensorFlow federated: machine learning on decentralized data, (2020).

75. S. Caldas et al., Leaf: A benchmark for federated settings, arXiv preprint arXiv:1812.01097, 2018.

76. Mothukuri V, Parizi R M, Pouriyeh S, Huang Y, Dehghantanha A, Srivastava G. A survey on security and privacy of federated learning. Future Generation Computer Systems, 2021, 115: 619–640

77. .Fung C, Yoon C J, Beschastnikh I. Mitigating sybils in federated learning poisoning. 2018, arXiv preprint arXiv: 1808.04866

78. Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan H B, Patel S, Ramage D, Segal A, Seth K. Practical secure aggregation for privacypreserving machine learning. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017,1175−1191

79. Zhao Y, Li M, Lai L, Suda N, Civin D, Chandra V. Federated learning with non-iid data. 2018, arXiv preprint arXiv: 1806.00582

80. 19.Li T, Sahu A K, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. IEEE Signal Processing Magazine, 2020, 37(3): 50–60

81. Ramaswamy S, Mathews R, Rao K, Beaufays F. Federated learning for emoji prediction in a mobile keyboard. 2019, arXiv preprint arXiv:1906.04329

82. Fallah A, Mokhtari A, Ozdaglar A. Personalized federated learning with theoretical guarantees: a modelagnostic meta-learning approach. Advances in Neural Information Processing Systems, 2020: 33

83. Ye D, Yu R, Pan M, Han Z. Federated learning in vehicular edge computing: a selective model aggregation approach. IEEE Access,2020, 8: 23920–23935

84. Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y. Federated learning for data privacy preservation in vehicular cyber-physical systems. IEEE Network, 2020, 34(3): 50–56

85. 12.Zhou C, Fu A, Yu S, Yang W, Wang H, Zhang Y. Privacy-preserving federated learning in fog computing. IEEE Internet of Things Journal, 2020, 7(11): 10782–10793