When Federated Learning Meets Vision: An Outlook on Opportunities and Challenges

Ahsan Raza Khan[1], Ahmed Zoha[1(✉)], Lina Mohjazi[1], Hasan Sajid[2], Qammar Abbasi[1], and Muhammad Ali Imran[1]

[1] James Watt School of Engineering, University of Glasgow, Glasgow G12 8QQ, UK
ahmed.zoha@glasgow.ac.uk

[2] Department of Robotics and Artificial Intelligence, SMME National University of Sciences and Technology (NUST), Islamabad, Pakistan

Abstract. The mass adoption of Internet of Things (IoT) devices, and smartphones has given rise to the era of big data and opened up an opportunity to derive data-driven insights. This data deluge drives the need for privacy-aware data computations. In this paper, we highlight the use of an emerging learning paradigm known as federated learning (FL) for vision-aided applications, since it is a privacy preservation mechanism by design. Furthermore, we outline the opportunities, challenges, and future research direction for the FL enabled vision applications.

Keywords: Federated Learning · Vision analytics · Edge computing · Decentralized data · Internet-of-Things · Collaborative AI

1 Introduction

According to international data corporation, there will be more then 80 billion devices (IoT sensors, smartphones, wearable sensors) connected to wireless networks by end of 2025. These devices will generate approximately 163 zeta bytes of data globally, which is 10 times of data generated in year 2016 [1,2]. The adoption of these devices are fueled by the advancements in wireless communications especially 5G technology. This overwhelming availability of data, advancement in deep learning, and unprecedented connectivity speeds offered by 5G will enable near real-time response for artificial intelligence (AI) driven applications.

The large-scale model training involves many stakeholders and entails many risks, which includes user privacy, data sovereignty, and data protection laws. The two common security attacks on a machine learning (ML) model are the poisoning attack (training phase) [3], and the evasion attack (inference phase) [4]. In the poisoning attack, the malicious user internally corrupts the training data, whereas in the evasion attack, the model accuracy can be manipulated by injecting adversarial samples. Therefore, different governments have introduced data

M. Ur Rehman and A. Zoha (Eds.): BODYNETS 2021, LNICST 420, pp. 308–319, 2022.
https://doi.org/10.1007/978-3-030-95593-9_23

protection regulations to ensure user privacy. To overcome this challenge, existing solutions are equipped with various privacy preserving techniques including differential privacy and modern cryptography techniques [5].

In recent times, differential privacy, coupled with powerful and advance wireless communications inspired many researchers to utilize the relevant data for many emerging AI driven applications [6,7]. However, the conventional cloud-centric model training approach requires transferring a large amount of raw data from the edge node to third-party servers. This however, has several limitations including:

- Data is privacy sensitive and highly protected under the legislation by General Data Protection Regulation (GDPR) [8].
- Latency issues incurred due to long propagation delays which are not acceptable in time-sensitive applications like smart healthcare, and self-driving cars [9].
- Inefficient bandwidth usage, higher communication and storage cost which also results in substantial network footprints.

This leads to the emergence of a new learning paradigm, termed as federated learning (FL) [10], which aims to bring computations to edge devices without compromising their privacy. Google being the pioneer, makes extensive use of FL algorithms to improvise their services like Gboard and next word prediction [11].

Though, FL was initially introduced with special emphasis on edge device and smartphone applications, but the combination of FL with IoT sensors and powerful AI tools has numerous applications in industry 4.0, digital health cares, smart cities, smart buildings, pharmaceutical drug discovered, video surveillance, digital imaging, virtual or augmented reality (VR/AR), and self-driving cars [12]. For instance, vision processing is an emerging technology, especially for healthcare and smart city applications. The vision sensors generate a large amount of data and it is challenging for the current wireless network architecture to process this data for time-sensitive applications. The key bottleneck is the communication cost and unprecedented propagation delays caused by the network congestion [1]. The 5G connectivity coupled with FL, is enabling a plethora of vision-aided applications, especially in smart healthcare, live traffic monitoring, and incident management [13]. The majority of these applications are privacy sensitive and latency intolerant. Therefore, the prospect of 5G connectivity and privacy by design of FL is envisioned to be a promising solution for vision-aided applications. Vision processing enabled by FL is an emerging field, therefore, it is very difficult to cover all related aspects. To this effect, in this article, we will discuss some of the possible verticals, system architecture, challenges, and future research directions of vision-aided applications.

The rest of the paper is organized as follows. Section 2 provides a brief overview of FL. Section 3 covers the possible vision-aided applications and review some of the use cases, whereas in Sect. 4, the detail of challenges and future research directions will be discussed. Finally, Sect. 5 will concludes the paper.

2 Preliminaries and Overview

FL is an algorithmic solution for collaborative model training with the help of many clients (smart phones, IoT sensors, and organizations) orchestrated by the centralized server, which keeps the training data decentralized [10]. It embodies the principle of relevant data collection and has the privacy by design. The concept of FL was initially introduced with special focus on smartphone and edge device applications, however, due to its decentralized nature of model training, it is also gaining popularity in other fields [12]. Therefore, keeping the common abstractions of different applications in mind, FL can be categorized based on the scale of federation, data partitioning, and privacy mechanism as shown in Fig. 1 [14].

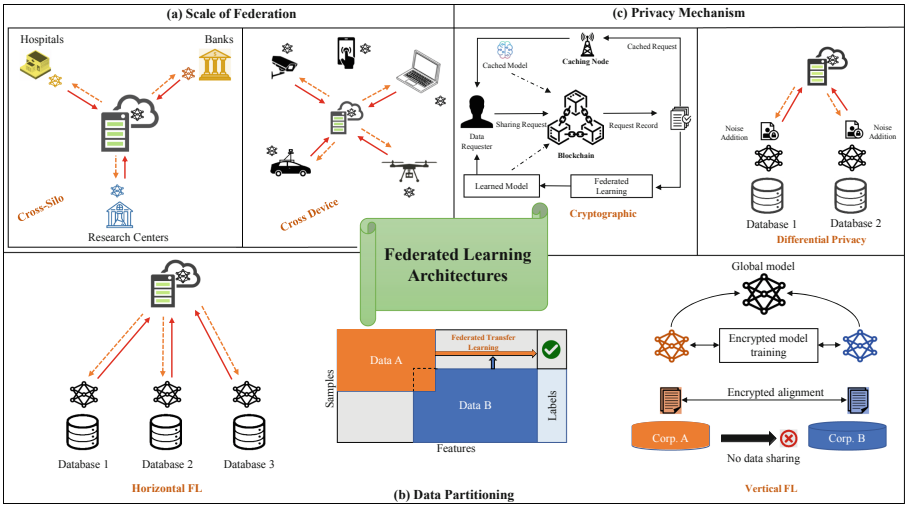

Fig. 1. Overview of different FL architectures.

2.1 Scale of Federation

The scale of federation is highly dependent on the number of edge nodes participating in the training process. When the clients in the training process are big organizations (hospitals, banks, and government institutions etc.), the number of participants will be small and this setting is called as cross-silo FL as shown in Fig. 1(a). Conversely, in cross-device FL settings, large number of users (smartphones, wearable sensors, and IoT sensors) participate in the training of a global model on a highly decentralized data-set [14]. The typical examples of cross-device and cross-silo FL are google Gboard and Nvida Clara for brain tumor segmentation respectively [10,15]. In vision-aided applications, the scale of federation is highly dependent on the nature of data and user's intent.

For instance, in brain tumor segmentation, the data was stored on central servers placed in different geographical locations, and as a result, the cross-silo mechanism is used for model training. In the smart cities video surveillance scenario, anomaly detection to identify the unusual activity in the environment is one of the of the examples. In this case, the cross-device mechanism may be used because this setting involves a large number of vision sensors placed in different locations.

2.2 Data Partitioning

FL is extremely useful in collaborative training where the data is distributed among a large number of users. In the era of digitization and big data, every click of user is captured to derive useful statistical information which may belong to similar or different application domains. Therefore, data partitioning plays a key role in FL where it is broadly divided in horizontal, vertical and transfer learning [14]. In horizontal FL, participants have similar features at different instances and vary in terms of data samples, whereas in vertical FL, common data of unrelated domains is used for model training. In vertical FL, users can have similar data but differ in terms of features. The classical example of horizontal FL is Google Gboard with the assumption of honest consumers and secure centralized server for global model training [10]. On the other hand, a real-world use case for vertical FL may be a scenario where the credit card sales team of a bank train its ML model by using the information of online shopping. In this case, only common users of the bank and e-commerce website will participate in the training process. With this liaising of secure information exchange, banks can improve their credit services and provide incentives to active customers [16].

In transfer FL approach, a pre-trained model is used on a similar datasets to solve a completely new problem set. The real-time example of transfer FL could be similar to vertical FL with small modifications. In this approach, the condition of similar users with matching data for model training can be relaxed to create a diverse system to serve individual customers [17]. It is a personalized model training for individual users to exploit the better generalization properties of global model which can be achieved by either data interpolation, model interpolation, and user clustering [18].

3 Vision-Aided Applications Enabled by FL

In recent times, vision processing has many practical applications in healthcare, smart transportation systems, video surveillance, and VR/AR. Conventional model training relies on server-led training solutions, however, video data is not only privacy sensitive but also incurs large communication cost as well [1]. FL mechanisms, on the other hand, are privacy-aware by design and significantly reduce the communication cost by exploiting the edge processing capabilities. Therefore, it is a very challenging task to build vision-aided solutions in a centralized server-led model. In recent times, FL is an exciting solution for

decentralized model training, which is gaining attention in both academia and industry. As a result of that, many FL enabled applications have surfaced. An overview of vision-aided applications enabled by FL for smart healthcare (cross-silo), smart transportation system and smart homes is presented in Fig. 2. It is very difficult to cover the entire liaison of FL applications, therefore, in this section we will focus on some of the vision-aided applications.

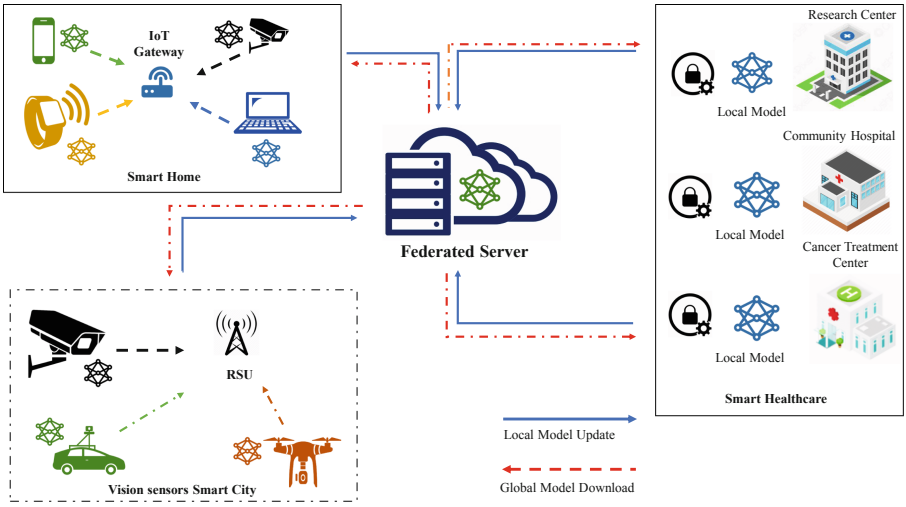

Fig. 2. Overview of vision-aided applications enabled by FL.

3.1 Smart Heath-Care System

In healthcare systems, data-driven ML is a promising approach to develop a robust model for learning features from large curated data for knowledge discovery. Even in the age of big data and advanced AI, the existing medical data is not yet fully exploited for model training due to privacy [19]. Most of the data is stored in secured locations i.e. a data island with restricted access. Furthermore, collecting, curating, and maintaining good quality data is both time consuming and expensive. Therefore, to improve the quality of health-care, collaborative learning without data share is a need of the hour and this platform is provided by FL. It is a promising solution to improve the healthcare data analytic, especially bio-medical imagining. FL can be applied on various domains of health-care but the key application areas involving vision data analytics are:

- Magnetic resonance imaging (MRI) to find neurological disease or disorders.
- Brain tumor segmentation.
- Human emotion detection to identify the mental health of patients.
- Cancer cell detection.

In [20], a FL framework is proposed to analyze the brain images to investigate neurological disorders. This framework used both synthetic as well as the real dataset, showing the potential of medical imaging in future applications. The authors in [21] presented a deep learning model for brain tumor segmentation using FL. In this study, multi-institutional collaboration is used which achieved the accuracy of 99% without sharing any data. Similarly, the authors in [22] exploited the client-server architecture of FL to train a differential privacy preserved deep neural network for brain tumor segmentation. In this study, the results show that there is a trade-off between privacy protection and model performance. In [23], a human emotion monitoring system is proposed using facial expression and speech signals to create an emotion index, which is used to find the mental health of individuals. Using FL, the proposed method showed promising results by detecting the depression of individuals without compromising the users' privacy.

3.2 Smart Homes

In smart homes, safety and security is highly dependent on vision processing solutions. Unfortunately, it is very difficult to deploy these solutions due to the privacy, latency, and high cost of video transmission. By addressing the privacy concerns effectively, real-time video analytic has many applications in smart homes. For instance, the combination of wearable sensors and real-time activity inference on indoor vision sensor feeds can help in fall detection and trigger corrective measures [24,25]. Similarly, for smart home safety, an alert can be triggered by visual instance detection which can identify the possible threat i.e., fire hazard. In [26], a visual object detection model $FedVision$ is presented which can be used to develop vision-aided solutions for safety monitoring in smart homes, cities, or industries. In this work, horizontal FL architecture is exploited to train the ML using the image data owned by a different organization to develop a warning mechanism for safety hazards. The experimental results showed improvement in operational efficiency, data privacy, and reduced cost.

3.3 Smart Cities

FL has a huge potential of effectively managing the assets, resources and services of smart cities using vision data analytics collected by vision sensors (smartphone cameras, CCTV, and dash-cams) [27]. With the challenge of privacy and high cost, cloud-centric approach also involves long propagation delay and incurs unacceptable latency for time sensitive applications like traffic and emergency management, self driving cars, disaster management [12]. For example, in smart transportation systems, a fleet of autonomous cars may need an up-to-date information of traffic, pedestrian behavior, or unusual incident (accident) to safely operate. Similarly, the video captured from individual smartphones or dash-cam can provide the live street view, which can be used for delivering the information of hospitals, popular restaurants, or providing insights on real-time behaviour of pedestrians and fellow drivers. However, building accurate models in these

scenarios will be very difficult due to the privacy and limited connectivity of each device. As a result, this can potentially impede the development of new technologies for smart cities [12,13]. Therefore, to reduce the transmission cost, and latency, FL can be used to locally process the information and only send the model parameter updates to the cloud. Using the FL paradigm, the following are the application domains of vision-aided solutions for smart cities.

- Smart transportation systems for real-time traffic management and navigation, incident detection, and automatic license plate/tag recognition.
- Self driving cars (automatic driving management and driver assistance).
- Safety and security of public places using the CCTV images and videos.
- Drone video surveillance for crowd management on special events.
- Natural disaster management using satellite imagery and drone footage.

In [28], a FL framework is proposed using unlabelled data samples at each user participating in the training process for two different application domains. The authors have demonstrated the application of FL and obtained promising results in natural disasters analysis and waste classification. Similarly, vision-aided applications enabled by FL also have a huge potential to improve the model training in some other domains like VR/AR, gaming, agriculture and smart industries, etc. The details of the used cases along with the area of applications are given in Table 1.

Table 1. Summary of vision aided applications enabled by FL.

Ref	Domains	Area of application	FL approach
[20]	Health-care	Neurological disorder	Cross-silo/Horizontal
[21]	Health-care	Brain tumor segmentation	Cross-silo/Horizontal
[22]	Health-care	Brain tumor segmentation	Cross-silo
[23]	Health-care	Human emotion detection	Cross-device/Horizontal
[26]	Smart homes	Visual object detection	Cross-device
[28]	Smart city	Disaster and waste classification	Cross-silo

4 Challenges and Future Research

FL is an emerging yet very effective and innovative learning paradigm for collaborative model training. Despite of recent research efforts to address the core challenges, FL is still prone to many limitations, especially in vision-aided application that hinder it to be adopted in different domains. In this context, we will discuss some of the challenges and future research directions.

4.1 Privacy and Security

In vision-aided applications, ML models are trained using highly sensitive data. Although, data never leaves the edge device during the training process, it is worth mentioning that FL does not address all the potential privacy issues. For instance, the FL trained model may indirectly leak some information to a third party user by model inversion, gradient analysis, or adversarial attacks [29]. Therefore, counter measures like adding noise, adding differential privacy is needed in cross-device architecture [12,14].

Level of trust among the participants in the training process is also a very big challenge in FL applications. In cross-silo structures, the clients are usually trustworthy and bounded by collaborative agreements, which reduces the trust deficit. As a result, there is a less possibility of privacy breach, which can help to reduce sophisticated counter protective measures [30]. However, in the cross-device architecture, the training process is done on a highly distributed dataset and it is almost impossible to enforce collaborative agreement. Therefore, trust deficit is a very big problem among the participants, and it is necessary to have some security strategies to ensure security and protect the end-user interests [19]. Similarly, privacy vs. performance trade-off is also a huge challenge in decentralized training, because it impacts the accuracy of final model [7].

4.2 Data Heterogeneity

The data captured by vision sensors is highly diverse, since it is collected by devices having different computational, storage, and network capabilities. For instance, an image or video captured by a smartphone or a dash-cam may have different pixel qualities [7]. Similarly, medical imaging data may also have distinct features and dimensions due to acquisition differences, quality and brand of the device, and local demographic bias [19,22]. Therefore, this non-identically distributed data poses a substantial challenge, which leads to the failure of a FL enabled solution under specific conditions. Data heterogeneity also leads to a situation where there is a conflict in the optimal solution and demands a sophisticated method to reach a global shared model. Therefore, data heterogeneity is still an open research problem and needs attention based on specific applications.

4.3 Asynchronous Aggregation Mechanism

Communication architecture for model aggregation is also huge challenge and is currently an active area of research. In cross-device model training, each device has different storage, computation and communication capabilities. Furthermore, device dropout is also very common in the training process due to connective and energy constraint [30]. These system level characteristics pose a critical challenge in model aggregation process. The traditional FedAvg algorithm uses the synchronous model aggregation mechanism, thus prone to the straggler effect in which FL server waits for all devices to complete their local training for global model update as shown in Fig. 3(a). This aggregation method slows down the

training process as it depends on the slowest device in the network. Furthermore, this mechanism does not account for a user who joins the training process halfway. On the other hand, asynchronous aggregation updates the global model as it receives the local update Fig. 3(b). One of the advantage of using asynchronous aggregation mechanism is its ability to deal with the straggler effect.

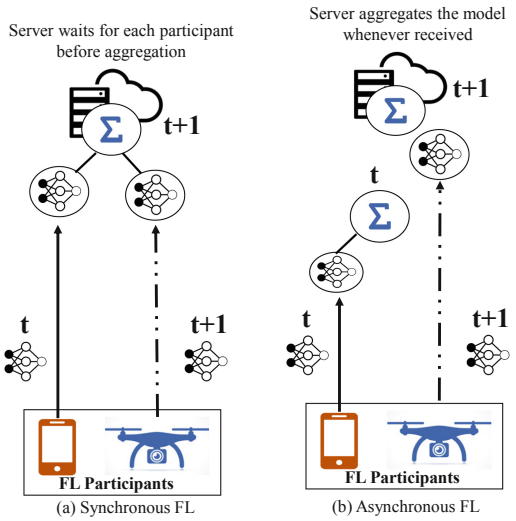

Fig. 3. Comparison of synchronous and asynchronous communication mechanism for FL.

4.4 Scale of Federation

The scale of federation is highly dependent on the number of participants in the training process for a specific application. For example, in health-care applications, cross-silo architecture is usually adopted for model training, where the edge nodes are hospitals and government institutes [19,30]. The participants in the training process are trust-worthy and equipped with secure communications, powerful computational resource. This is quite a straight forward training process and each client is bounded by the collaborative agreement. However, in smart city, cross-device architecture can be used for applications like transportation system and self-driving cars. The fully decentralized model training has many challenges including communication and propagation delays, trust deficit among client, model convergence and aggregation, and agreement of optimal solution [19]. This aspect of vision-aided application is unexplored.

4.5 Accountability and Incentive Mechanism

Data quality in ML-driven applications is essential because the performance of the system is highly dependent on the data. In FL model training, the data

quality has more significance. In non-trusted federation, the accountability of clients is very important to improve the performance of the model. This information can be used to develop a revenue model to give incentives and encourage the participants with relevant data to participate in the model training and improve the global model accuracy.

5 Conclusions

Data-driven solutions have led to a wide range of innovations, especially in the domain of vision-based applications and services. However, a lot of intelligence still remains untapped because of inaccessibility of user-centric information due to privacy challenges. Federated learning mechanism has led us to an exciting research paradigm that allows us to collect and analyze the massive amount of information without compromising on privacy and network resources. In this paper, we have provided an outlook on FL-enabled vision-aided applications. Furthermore, we have set the scene for vision applications in the era of 5G connectivity through FL paradigm and highlighted a number of fundamental challenges and future research directions.

References

1. Lim, W.Y.B., et al.: Federated learning in mobile edge networks: a comprehensive survey. IEEE Commun. Surv. Tutor. **22**(3), 2031–2063 (2020)
2. Zhu, G., Liu, D., Du, Y., You, C., Zhang, J., Huang, K.: Toward an intelligent edge: wireless communication meets machine learning. IEEE Commun. Mag. **58**(1), 19–25 (2021)
3. Muñoz-González, L., et al.: Towards poisoning of deep learning algorithms with back-gradient optimization. In: Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security, AISec 2017. ACM Press (2017)
4. Nasr, M., Shokri, R., Houmansadr, A.: Comprehensive privacy analysis of deep learning: passive and active white-box inference attacks against centralized and federated learning. In: IEEE Symposium on Security and Privacy, pp. 739–753 (2019)
5. Bae, H., et al.: Security and privacy issues in deep learning. arXiv preprint arXiv:1807.11655
6. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: challenges, methods, and future directions. IEEE Signal Process. Mag. **37**(3), 50–60 (2020)
7. Li, P., et al.: Multi-key privacy-preserving deep learning in cloud computing. Futur. Gener. Comput. Syst. **74**, 76–85 (2017)
8. Custers, B., Sears, A.M., Dechesne, F., Georgieva, I., Tani, T., van der Hof, S.: EU Personal Data Protection in Policy and Practice. TMC Asser Press, The Hague (2019)
9. Yang, K., Jiang, T., Shi, Y., Ding, Z.: Federated learning via over-the-air computation. IEEE Trans. Wireless Commun. **19**(3), 2022–2035 (2020)
10. McMahan, H.B., Moore, E., Ramage, D., Arcas, B.A.: Federated learning of deep networks using model averaging. arXiv preprint arXiv:1602.05629 (2016)

11. McMahan, B., Moore, E., Ramage, D., Hampson, S., Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial Intelligence and Statistics, pp. 1273–1282 (2017)
12. Aledhari, M., Razzak, R., Parizi, R.M., Saeed, F.: Federated learning: a survey on enabling technologies, protocols, and applications. IEEE Access **8**, 140699–140725 (2020)
13. Deng, Y., Han, T., Ansari, N.: FedVision: federated video analytics with edge computing. IEEE Open J. Comput. Soc. **1**, 62–72 (2021)
14. Mothukuri, V., Parizi, R.M., Pouriyeh, S., Huang, Y., Dehghantanha, A., Srivastava, G.: A survey on security and privacy of federated learning. Futur. Gener. Comput. Syst. **115**, 619–640 (2021)
15. https://resources.nvidia.com/en-us-federated-learning/bvu-ea6hc0k?ncid=pa-srch-goog-84545
16. Yang, S., Ren, B., Zhou, X., Liu, L.: Parallel distributed logistic regression for vertical federated learning without third-party coordinator. arXiv preprint arXiv:1911.09824 (2019)
17. Chen, Y., Qin, X., Wang, J., Yu, C., Gao, W.: FedHealth: a federated transfer learning framework for wearable healthcare. IEEE Intell. Syst. **35**(4), 83–93 (2020)
18. Mansour, Y., Mohri, M., Ro, J., Suresh, A.T.: Three approaches for personalization with applications to federated learning. arXiv preprint arXiv:2002.10619 (2020)
19. Rieke, N., et al.: The future of digital health with federated learning. NPJ Digit. Med. **3**(1), 1–7 (2020)
20. Silva, S., Gutman, B.A., Romero, E., Thompson, P.M., Altmann, A., Lorenzi, M.: Federated learning in distributed medical databases: meta-analysis of large-scale subcortical brain data. In: IEEE 16th International Symposium on Biomedical Imaging, pp. 270–274 (2019)
21. Sheller, M.J., Reina, G.A., Edwards, B., Martin, J., Bakas, S.: Multi-institutional deep learning modeling without sharing patient data: a feasibility study on brain tumor segmentation. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11383, pp. 92–104. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11723-8_9
22. Li, W., et al.: Privacy-preserving federated brain tumour segmentation. In: Suk, H.-I., Liu, M., Yan, P., Lian, C. (eds.) MLMI 2019. LNCS, vol. 11861, pp. 133–141. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32692-0_16
23. Chhikara, P., Singh, P., Tekchandani, R., Kumar, N., Guizani, M.: Federated learning meets human emotions: a decentralized framework for human-computer interaction for IoT applications. IEEE Internet Things J. **8**(8), 6949–6962 (2021)
24. Wu, Q., Chen, X., Zhou, Z., Zhang, J.: FedHome: cloud-edge based personalized federated learning for in-home health monitoring. IEEE Trans. Mob. Comput. (2020)
25. Sozinov, K., Vlassov, V., Girdzijauskas, S.: Human activity recognition using federated learning. In: IEEE International Conference on Parallel & Distributed Processing with Applications, pp. 1103–1111 (2018)
26. Liu, Y., et al.: FedVision: an online visual object detection platform powered by federated learning. In: Proceedings of the Conference on Artificial Intelligence, vol. 34, no. 08, pp. 13172–13179 (2020)
27. Zheng, Z., Zhou, Y., Sun, Y., Wang, Z., Liu, B., Li, K.: Federated learning in smart cities: a comprehensive survey. arXiv preprint arXiv:2102.01375 (2021)
28. Ahmed, L., Ahmad, K., Said, N., Qolomany, B., Qadir, J., Al-Fuqaha, A.: Active learning based federated learning for waste and natural disaster image classification. IEEE Access **8**, 208518–208531 (2020)

29. Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., Qi, H.: Beyond inferring class representatives: user-level privacy leakage from federated learning. In: IEEE INFOCOM IEEE Conference on Computer Communications, pp. 2512–2520 (2019)
30. Kairouz, P., et al.: Advances and open problems in federated learning. arXiv preprint arXiv:1912.04977 (2019)