A Survey on Federated Learning for Intelligent Healthcare Systems

[1]Mrs. Deepthi S, [2]Dr.R.Chinnayan

Assistant Professor, School of Computer Science & Engineering, Presidency University,Bangalore, India

Professor, Department of Computer Science & Engineering, Presidency University, Bangalore, India

DOI: https://doi.org/10.5281/zenodo.10040680

Published Date: 25-October-2023

Abstract: **The accelerated growth of Artificial Intelligence (AI) has greatly influenced the healthcare industry, offering significant advancements in smart healthcare systems. However, the lack of standards, legal regulations, and is difficult to meet ethical standards for patient information privacy. The utilization of large quantities of user data for training machine learning models has shown promising results. Nonetheless, two major obstacles persist: the fragmented nature of user data, hindering aggregation without compromising privacy, and the failure of cloud-based models to personalize healthcare. To address these issues, Federated Learning (FL) has emerged as a solution, leveraging privacy-preserving algorithms to overcome data atomization concerns. Furthermore, integrating FL with technologies like blockchain and edge computing can enhance security and computational efficiency.**

This paper presents an overview of FL architectures, comparing many kinds of federated learning frameworks and distributed machine learning algorithms. It explores the limitations of current smart healthcare systems and highlights how FL can overcome these challenges. The study investigates different FL architectures and classification models, showcasing their potential application in healthcare.

Furthermore, it analyses the advantages of FL in medical settings, emphasizing privacy preservation and improved data management. The paper also assesses the security risks associated with healthcare applications and proposes ways to mitigate them. The research findings aim to help both academia and industry understand the competitive advantage offered by advanced privacy-preserving federated learning systems in the field of healthcare data.

Keywords: **Artificial Intelligence (AI), Federated Learning, Privacy Preservation, Data Management, Security Risks.**

I. INTRODUCTION

Traditional machine learning and deep learning technique approach necessitates the transfer of data from multiple devices, individuals, or institutions, resulting in high computational expenses due to the large dataset requirements. Privacy preservation becomes a significant challenge when dealing with sensitive data like medical information. Centralized databases are prone to hacking attacks, significantly elevating the risk of data breaches.

Google introduced federated learning (FL) as a method for lowering the cost of computing by utilizing the processing power of mobile devices [1-3]. FL operates by conducting training at the individual client level, where each client's local weights are transmitted to the server. The server aggregates the updated local weights and computes new global weights. The clients then download the global weights from the server and continue the training process. FL was initially employed in mobile apps [4-6] and has since been extensively researched and enhanced across various domains [7-11]. Significant attention has been given to exploring FL in relation to data heterogeneity [1,12], robust optimization [13-17], and security measures such as differential privacy and secure multiparty computation [13,18,19]. The medical field has also witnessed research on FL, specifically utilizing EMR and brain tumor data [20-22]. Federated learning (FL) is advantageous for medical data analysis, considering the sensitivity of such personal information. Traditionally, deidentification methods have been employed to protect patient privacy [23-25]. However, these methods typically necessitate data centralization which increases the risk of data breaches.

The Health Insurance Portability and Accountability Act (HIPAA) in the United States provides specific deidentification guidance, outlining 18 types of protected health information that should be removed [26]. However, many researchers and advocates argue that this guidance should be revised to enhance privacy protection [27]. In contrast, FL eliminates the need for raw data centralization. Even FL developers do not have access to the raw data, ensuring enhanced privacy and resolving issues related to privacy protection or deidentification when using clinical data.

II. LITERATURE REVIEWS

In the context of smart healthcare systems, training machine learning models necessitates access to large and diverse datasets, which poses a significant challenge. Furthermore, the availability of data and the need to protect individual privacy pose major obstacles to traditional centralized machine learning approaches in healthcare.

2.1 Distributed Machine Learning

Distributed machine learning model parallelism is a technique used to train large-scale ML models by distributing the computational workload across multiple machines or devices. It is useful when working with models that are excessively large to fit in an individual machine's memory or whenever the training data is too big for a single device to handle.

Distributed machine learning refers to a system where machine learning tasks are performed across multiple nodes or computing devices which brings several benefits, including reduced training time, the ability to train larger models, and improved scalability. However, it also introduces challenges such as increased communication overhead, synchronization issues, and the need for efficient distributed training. By distributing the computational workload across multiple machines or nodes, these algorithms leverage parallel processing capabilities to overcome the limitations of single-machine approaches. Table 1and Table 2 compares and contrasts various Distributed and Federated learning architectures.

Table 1. Comparison of DML and FL architectures.

	Distributed Machine Learning (DML)	Federated Learning (FL)
Data Centralization	Centralized data storage and processing	Decentralized data on individual devices/clients
Privacy and Security	Risk of data breaches and unauthorized access to centralized data	Privacy-preserving, data remains on client devices
Communication	Communication between central server and distributed nodes	Minimal communication, only model updates are exchanged
Scalability	Scalable with addition of more machines/nodes	Scalable to a large number of client devices participating
Data Availability	Requires centralized collection of data	Works with fragmented or isolated datasets
Computational Cost	High computational cost due to data transfer and processing	Lower computational cost, training occurs on client devices
Model Updates	Centralized model updates based on aggregated data	Decentralized model updates based on individual client training
Flexibility	Less flexible due to centralization	More flexible as data remains with clients

Table 2. Analyses of DML and FL architectures.

	Distributed Machine Learning (DML)	Federated Learning (FL)
Architecture	Centralized data storage and processing	Decentralized data on individual devices/ clients
Advantages	Scalable with addition of more machines /nodes	Privacy-preserving, data remains on client devices
	Can handle large data -sets and complex models	- Utilizes distributed data, enables broader collaboration
	Provides centralized model updates for better coordination	- Reduces communication and computational costs
	Access to diverse and representative datasets	- Allows for personalized models and edge device capabilities

Limitations	- Privacy concerns with centralized data storage	- Relies on reliable and well-connected client devices
	- Higher communication and computational costs	- Limited to the computational capabilities of individual clients
	- Requires data availability and collection for centralization	- May result in biased or non-representative global model
	- May face challenges in heterogeneous data scenarios	- Potential risk of model poisoning or Byzantine behavior

2.2 Federated Learning

A networked machine learning method called federated learning allows several healthcare systems to jointly train a single model while maintaining autonomous and secure data management. Here's a simplified diagram (Fig.1) illustrating the components and flow of a federated learning model for healthcare systems:

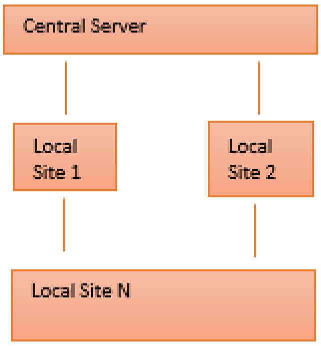

Fig.1 Simple Federated Learning Model for Healthcare Systems

1. Central Server: This is the central coordination point that manages the federated learning process. It coordinates model updates and aggregates the results from the local sites. The server typically hosts the global model and communicates with the local sites.

2. Local Sites: These are the individual healthcare systems or institutions that participate in the federated learning process. Each local site has its own data and computational resources. They train the local models using their respective datasets while keeping the data locally.

3. Local Model Training: Each local site uses a neural network or another machine learning model appropriate for the current healthcare task to train the local model on the data that is locally accessible.

4. Model Update: Once the local models are trained, they send their updates to the central server. The updates usually consist of the model parameters or gradients computed during training.

5. Model Aggregation: After gathering the model updates, from the local sites, the central server performs an aggregation step to combine the updated values into a new global model. This aggregation could be as simple as averaging the model parameters or using more sophisticated techniques like federated averaging or secure multi-party computation.

6. Global Model Update: After the aggregation step, the central server updates the global model with the newly obtained parameters. The updated global model is then sent back to the local sites.

7. Iterative Training: The process of local model training, update submission, aggregation, and global model update is typically repeated iteratively. Each iteration allows the local models to learn from the collective knowledge of all participating sites while preserving data privacy.

Federated learning can be divided into three categories: Horizontal federated learning, Vertical federated learning, and Federated transfer learning, considering the data sample space distribution and different data feature spaces.

2.2.1 Horizontal federated learning (HFL)

Sample-based federated learning, also known as horizontal federated learning, involves sharing data horizontally among different parties. In this approach, the parties exchange their model updates with an aggregator or central server which collects these model updates and combines them to generate a global model. This combination involves merging or averaging the model weights obtained from different parties. By aggregating the model updates, the central server creates a new global model that benefits from the knowledge learned by each party while preserving the privacy of their data. This approach is particularly useful when the datasets across different parties have similar feature spaces but differ in the samples they contain. It allows for collaborative model training and leverages the collective knowledge from diverse datasets without the need to share the raw data (Fig.2).

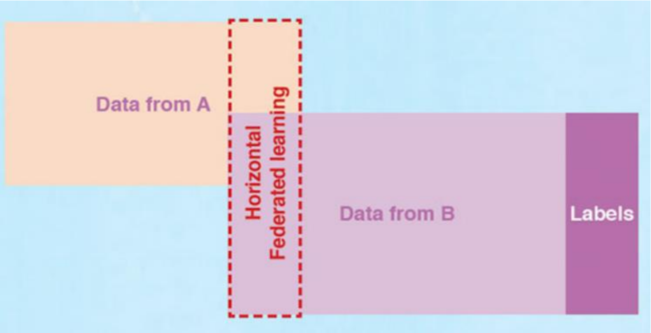

Fig.2 Horizontal Federated Learning

For example [28], The information about patients is kept in a database as groups in the blood bank laboratory. The database has features like Name, Age, and Blood group. Persons 1, 2, and 3 in Fig. 3 show sets of individuals whose data samples share Name, Age, and Blood Group characteristics. The Horizontal FL model is applied on each of these samples (person) separately. The local model will be trained on these samples. Finally, a global model will be generated by aggregating model weights.

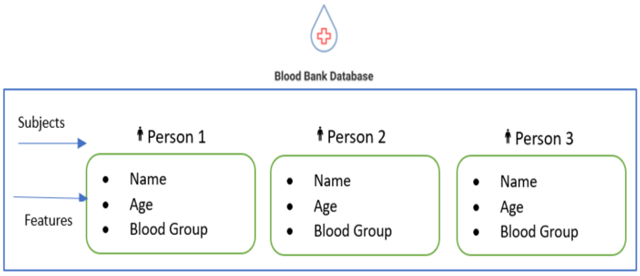

Fig.3 Horizontal FL Model

2.2.2. Vertical federated learning (VFL)

It is also known as feature-based federated learning, can be applicable to the cases where two data sets share the same sample ID space but may vary in feature space. This is used in the cases where data shared contains similar samples but different features. When multiple parties are involved and none of them have access to the whole set of features and labels, VFL is applicable. As a result, they are unable to locally train a model using their datasets where features are vertically aggregated. To jointly create a model using data from both parties, the training loss and gradients are computed in a privacy-preserving way (Fig.4).

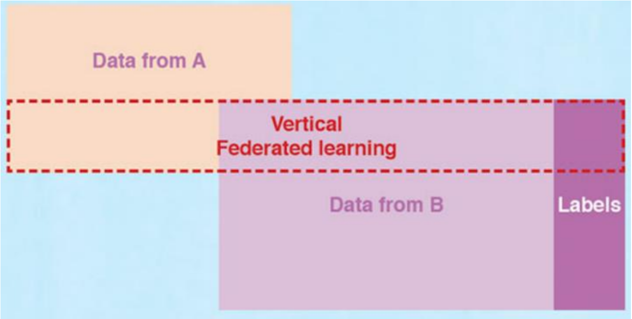

Fig.4 Vertical Federated Learning

In Fig. 4, In VFL[29], the first sample for the model of machine learning would be a data set from Person 1. The second dataset will contain the data samples from Person 2 from the blood bank database and the hospital database, respectively.

2.2.3. Federated transfer learning

It was created to address the issue of integrating the dispersed data and enhance statistical modelling when carrying out data federation. The model structure needs only minor adjustments for the federated transfer learning framework (Fig. 5), and the output is just as effective as non-privacy protecting transfer learning. As seen in Fig. 6, for instance, patients in hospitals and pharmacies in different cities are not all from the same population, and some patient characteristics do not match very well [30]. Federated transfer learning doesn't rely on any specifications, such a shared feature space or sample area. Instead, data federation promotes transfer learning by providing answers for the entire population and feature space.

FedHealth is one such approach that applies the concept of distributed transfer learning to intelligent wearable medical technology[31]. The method generates personalised models by transfer learning while performing data aggregation through federated learning, all while maintaining the confidentiality and security of the model and data.

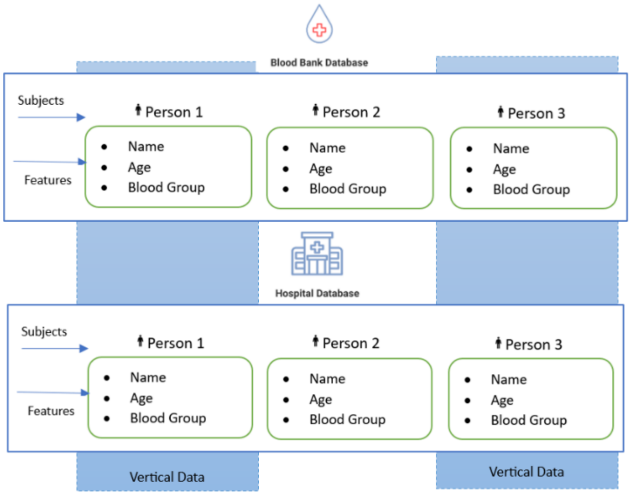

Fig.5 Vertical FL Model

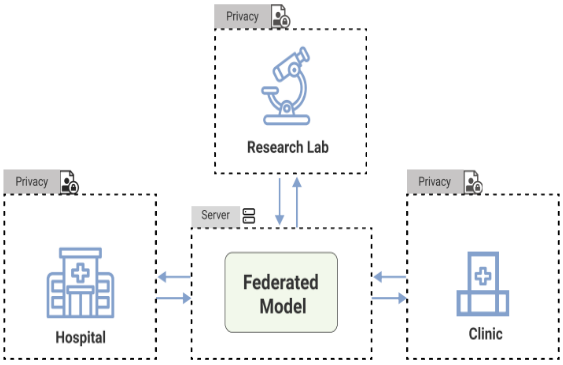

Fig.6 Transfer learning in federated model.

Limitations of Current Smart Healthcare Systems

a. Privacy Issues: The privacy issues associated with a centralized AI-based smart healthcare system are substantial, considering the fact that cloud servers can provide efficient data training and analytics with powerful compute capabilities.

b. Scarcity of Datasets at Medical Sites: Healthcare systems may not have enough data from a single medical facility (such as a clinical lab) to run the AI model, which can hinder the model from properly training on health data. Here, data analyses must be performed manually, which causes significant delays in data processing. [32] It is not simple to collect data from other sites to train the AI model because of institutional rules and growing user privacy concerns[33].

c. Insufficient Health Data Training Performance: Due to the imbalanced data features and the inadequate data volumes, training at a single healthcare location cannot achieve the required degree of accuracy, such as sickness classification accuracy. Techniques for data augmentation, such as generative adversarial networks, can be used to tackle these issues(GANs) [42], but it can still lack sufficient diversity to create a sizable dataset for effective data training.

d. Medical Data Training is costly: Because medical data are frequently huge (e.g., audio, pictures), AI-based smart healthcare systems experience high network delay and bandwidth when sending health data to the cloud for processing [34] which may lead to congestion. Medical device transmission power is also required throughout the offloading process, which presents new hardware and battery design problems.

2.3 The utilization of Federated Learning in the healthcare domain.

The largest problem with medical applications is the absence of publicly accessible multicentered and heterogeneous datasets. Federated learning solves this problem as well as the privacy and confidentiality concerns that arise [28]. Table 2 shows some of the observations drawn from recent healthcare applications where FL is used.

2.4 Advantages of FL in Smart Healthcare

By employing federated learning (Fig.7), healthcare systems can collaborate and improve their models collectively without directly sharing sensitive patient data. This strategy aids in addressing privacy issues while utilizing the collective expertise of several organizations to produce more precise and reliable models. Federated learning has the potential to bring several benefits to healthcare systems.

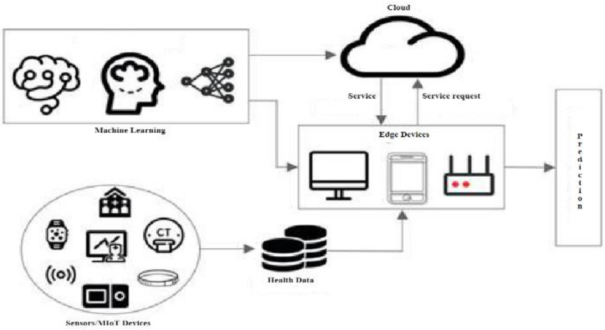

Fig.7 FL Based Smart Healthcare Systems

Here are some of the key advantages of FL:

1. *Privacy Preservation:* FL helps to maintain data privacy and confidentiality, as individual data remains local, and only model updates or aggregated insights are shared.

2. *Data Security:* By keeping data decentralized and performing computations locally, federated learning reduces the risk of data exposure during transmission or storage, enhancing overall data security.

3. *Large-Scale Data Access:* Federated learning enables healthcare organizations to pool their data resources while maintaining data ownership. It enables hospitals, clinics, and research institutions to collaborate and benefit from a diverse range of patient data, leading to more robust and generalizable models.

4. *Improved Model Performance:* Machine learning models are trained via federated learning, which taps on the combined intelligence of many participants. This can lead to improved model performance and generalization, as the model gains insights from a broader population without relying on a single data source.

5. *Reduced Data Bias*: Federated learning addresses this issue by training models on data from multiple sources, including diverse patient populations. It helps mitigate bias by ensuring that the resulting models are more representative and fairer across different subpopulations.

6. *Real-Time Learning*: Federated learning facilitates continuous learning from decentralized data sources. As new data becomes available at the edge devices or local servers, real-time learning capability enables healthcare systems to adapt quickly to evolving conditions, such as emerging diseases or changing patient profiles.

7. *Cost-Efficiency*: Since majority of computations are performed locally, FL results in significant cost savings, as the infrastructure and bandwidth requirements for centralized data storage and processing are reduced. Moreover, federated learning allows healthcare organizations to collaborate and share the development and maintenance costs of machine learning models, making it a cost-effective approach.

Table.3 Observations on FL based Intelligent HealthCare Systems.

Applied domain	Ref.	FL type	FL clients	FL aggregator	Key contributions	Limitations
Management of EHRs	[78]	HFL	Smart phones	Data server	FL system for forecasting patient hospitalizations	Simple approach lacking in-depth analyses
Management of EHRs	[79]	HFL	Hospitals	Data server	Differential privacy-based FL for federated EHRs training	Viability of the model in real-world FL implementations should be considered
Health Monitoring	[80]	VFL	Smart phones	Cloud server	Monitoring remote activity recognition with a personalized FL scheme.	There has been no consideration for secure aggregation in FL communications
Health Monitoring	[81]	FTL	Wearable devices	Cloud server	Personalized FL scheme that uses transfer learning to recognize human activity.	The communication costs and complexity of training have not been verified.
Health Monitoring	[82]	HFL	Mobile devices	Data centre	an approach based on FL for mood prediction on mobile.	In mobile FL, privacy and the use of training resources should be taken into account.
Medical Imaging	[83]	HFL	Medical sites	Data centre	a brain imaging MRI analysis framework built on FL.	The federated MRI training should be applied to practical components.
Medical Imaging	[84]	HFL	Hospitals	Data centre	An FL scheme to help with acute neurological symptom diagnosis.	Simulations must be made more often and simply.
Medical Imaging	[85]	HFL	Hospitals	Data centre	An FL-based method for medical imaging with differentiated privacy for safe, cooperative training.	No DL method comparison in FL simulation.
COVID-19 Detection	[86]	HFL	Medical institu-tions	Data centre	a COVID-19 screening method based on chest X-ray images using FL.	FL communications-related data loss has not been taken into account.
COVID-19 Detection	[87]	VFL	Hospitals	Aggregator	a federated deep learning technique to detect lung abnormalities in COVID19	There has been no analysis of training delay.
COVID-19 Detection	[88]	VFL	Medical institu-tions	Cloud server	Using international data from China, Italy, and Japan, a FL approach for COVID region segmentation in chest CT.	It should be possible to offer a theoretical study of FL communications and convergence.

Table.4 Performance of FL in recent medical applications.

Literature	Data Sets	Applications of healthcare	Algorithms	Results
[89]	EHR	COVID -19 patients' Mortality prediction	MLP and FL	Federal aggregation gives better results than centralized
[90]	Ultrasound images	Disease prediction	FL and ResNet -50	The efficiency of federated deep learning and traditional deep learning is equivalent.
[91]	Dermatology Atlas data set	Skin disease detection	FL and DualGAN	Strong security and robustness
[92]	Mental Health Dataset	Depression analysis	multiview federated learning framework	The reliability of depression prediction improved
[93]	Rehabilitation Cancer Data Sample Set	Cancer Diagnosis Model	FL with CNN	Achieved promising prediction results by protecting patients data.

Consider the following suggestions for cutting communication expenditures.

i. *Data Filtering and Compression:* Apply data filtering techniques before sending data between clients and the central server to weed out any redundant or unneeded information. Reduce the amount of the data being communicated by using data compression methods to further cut down on communication expenses.

ii. *Model Compression and Quantization:* By applying model compression techniques, it is possible to minimize the dimension of the models that are transferred between users and the main server, that help reduce the model's parameters and overall size without significantly affecting performance.

iii. *Local Model Training:* Before sending changes to the central server, run further local model training iterations on the client devices. This lessens the number of rounds of communication, which cuts down on the price of communicating updates back and forth.

iv. *Differential Privacy:* Introduce noise to the gradients or model updates before sharing them by using differential privacy methods. This helps safeguard the confidentiality of specific data samples while yet enabling the central server to gain knowledge from aggregated data. The demand for transmitting fine-grained data can be lessened to cut down on communication expenses.

v. *Selective Aggregation:* Selectively aggregate updates from a group of customers or in accordance with predetermined criteria rather than averaging updates from all clients. By limiting the quantity of data transferred during aggregation, this method lowers communication costs while retaining the global model's representativeness.

vi. *Communication-Efficient Algorithms:* Improved FL algorithms aim to decrease the amount of data exchanged or the amount of communications rounds required without compromising the quality of the learned model.

vii. *Edge Computing:* Utilize edge computing resources by running local model inference and training on edge servers or edge devices. With less frequent contact with a central server required, communication costs may drop dramatically.

viii. *Network Bandwidth Optimization:* Prioritize data transfers that are absolutely necessary, use data compression methods, and take use of communication overhead-reducing strategies like data batching or pipelining to maximize network capacity.

2.5 Issues and Challenges

The core challenges of Federated Learning are listed below:

- Communication efficiency throughout the federated network.
- Managing multiple systems in the same network.
- Data in federated networks has statistical heterogeneity.
- Concerns about privacy and methods to protect it.

Here are some of the key issues.

1. Data heterogeneity
2. Data privacy and security
3. Communication and bandwidth limitations
4. Imbalanced data distribution
5. Model performance and generalization
6. Governance and collaboration
7. Interoperability and standards

Addressing these challenges requires a combination of technical advancements, policy frameworks, and collaboration among various stakeholders. Continued research and development efforts are needed to overcome these obstacles and realize the full potential of federated learning in healthcare.

III. CONCLUSION AND PROSPECT

The latest literature on federated learning and the healthcare sector is reviewed in this study. Firstly, introducing importance of smart healthcare systems. The paper outlines the architecture of various FL models, compares various types of Distributed machine leaning and Federated learning architecture, the limitations of current smart healthcare systems and how FL can overcome the same. We also go through various federated learning architectures and its usage in healthcare, and analyze its benefits for use in medical applications. The paper also analyses various security risks faced by healthcare applications, concludes with a discussion of previous work done in the field of FL in smart healthcare applications and also outlines the limitations of the existing FL models.

The results of this study assist colleagues in academia and business in realising the competitive advantage of the most cutting-edge federal learning systems for healthcare data management that protect privacy.

REFERENCES

[1] Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, Dave Bacon, "Federated learning: strategies for improving communication efficiency". 30 Oct 2017 arXiv:1610.05492. https://doi.org/10.48550/arXiv.1610.05492.

[2] McMahan H, Moore E, Ramage D, Hampson S. Communication-efficient learning of deep networks from decentralized data. 26 Jan 2023 arXiv preprint 2016:1602.05629. https://doi.org/10.48550/arXiv.1602.05629

[3] Brendan H. "Federated Learning: Collaborative Machine Learning without Centralized Training Data", Google AI Blog. 2017. URL: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html [accessed 2020-10-13].

[4] Hard A, Rao K, Mathews R, Ramaswamy S, Beaufays F, Augenstein S. "Federated learning for mobile keyboard prediction", arXiv preprint 2018:1811.03604.

[5] Yang T, Andrew G, Eichner H, Sun H, Li W, Kong N. "Applied federated learning: Improving google keyboard query suggestions", arXiv preprint 2018:1812.02903.

[6] Chen M, Mathews R, Ouyang T, Beaufays F. "Federated learning of out-of-vocabulary words", arXiv preprint 2019:1903.10635.

[7] Rieke N, Hancox J, Li W, Milletarì F, Roth HR, Albarqouni S, et al. "The future of digital health with federated learning", NPJ Digit Med 2020;3:119.

[8] K. Kairouz and H. McMahan, "Advances and Open Problems in Federated Learning," arXiv preprint arXiv:1912.04977, 2019.

[9] Li T, Sahu AK, Talwalkar A, Smith V. "Federated Learning: Challenges, Methods, and Future Directions", IEEE Signal Process Mag 2020 May;37(3):50-60.

[10] Jie X, Benjamin S, Glicksberg, Chang S, Peter W, Jiang B, et al. "Federated learning for healthcare informatics", arXiv preprint 2019:1911.06270.

[11] Qinbin L, Zeyi W, Bingsheng H. "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection" arXiv preprint 2019:1907.09693.

[12] Zhao Y, Meng L, Liangzhen L, Naveen S, Damon C, Vikas C. "Federated Learning with Non-IID Data" arXiv preprint 2018:1806.00582.

[13] Li T, Sahu A, Zaheer M, Sanjabi M, Talwalkar A, Smith V. "Federated Optimization in Heterogeneous Networks" arXiv preprint 2018:1812.06127.

[14] Konečný J, McMahan H. "Federated Optimization: Distributed Machine Learning for On-Device Intelligence" arXiv preprint 2016:1610.02527.

[15] Wang Y. "Cooperative Machine Learning from Mobile Devices Masters Thesis" University of Alberta. 2017. URL: https://era.library.ualberta.ca/items/7d680f04-7987-45c5-b9cd-4fe43c87606f [accessed 2020-10-15]

[16] Huang L, Yin Y, Fu Z, Zhang S, Deng H, Liu D. "LoAdaBoost: loss-based AdaBoost federated machine learning with reduced computational complexity on IID and non-IID intensive care data" arXiv preprint 2018:1811.12629.

[17] Pillutla K, Kakade S, Harchaoui Z. “Robust Aggregation for Federated Learning”, arXiv preprint 2019:1912.13445.

[18] Abadi M, Chu A, Goodfellow I, McMahan H, Mironov I, Talwar K, et al. “Deep Learning with Differential Privacy”. 2016 Presented at: CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; 2016; Vienna.

[19] Pfohl Stephen R, Andrew M, Katherine H. “Federated and Differentially Private Learning for Electronic Health Records”. arXiv preprint 2019:1911.05861.

[20] Li W. “Privacy-Preserving Federated Brain Tumour Segmentation”, Springer, Cham; 2019 Presented at: Machine Learning in Medical Imaging. MLMI 2019; 2019; Shenzhen.

[21] Sheller MJ, Reina GA, Edwards B, Martin J, Bakas S. “Multi-Institutional Deep Learning Modeling Without Sharing Patient Data: A Feasibility Study on Brain Tumor Segmentation”, 2019 Presented at: International MICCAI Brain Lesion Workshop; 2018; Granada p. 92-104 URL: http://europepmc.org/abstract/MED/31231720.

[22] Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. “Federated learning of predictive models from federated Electronic Health Records”, Int J Med Inform 2018 Apr;112:59-67.

[23] El Emam K, Dankar FK, Issa R, Jonker E, Amyot D, Cogo E, et al. “A globally optimal k-anonymity method for the de-identification of health data.” J Am Med Inform Assoc 2009;16(5):670-682.

[24] Taira RK, Bui AA, Kangarloo H. “Identification of patient name references within medical documents using semantic sectional restrictions.” Proc AMIA Symp 2002:757-761.

[25] Thomas SM, Mamlin B, Schadow G, McDonald C. “A successful technique for removing names in pathology reports using an augmented search and replace method.” Proc AMIA Symp 2002:777-781.

[26] U.S. Department of Health & Human Services, "Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule," HHS.gov, Washington, DC, USA, 1996. [Online]. Available: https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html. [Accessed: Oct. 13, 2020].

[27] Nass SJ, Levit LA, Gostin LO. “Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research.” Washington, DC: National Academies Press; 2009.

[28] M. Joshi, A. Pal, and M. Sankarasubbu, "Federated Learning for Healthcare Domain - Pipeline, Applications and Challenges," Saama AI Research, India, October 2022.Available: https://doi.org/10.1145/3533708.

[29] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated Learning: Strategies for Improving Communication Efficiency," CoRR, abs/1610.05492, 2016. [Online]. Available: http://arxiv.org/abs/1610.05492.

[30] Ce Ju, Ruihui Zhao, Jichao Sun, Xiguang Wei, Bo Zhao, Yang Liu, Hongshan Li, Tianjian Chen, Xinwei Zhang, Dashan Gao, Ben Tan, Han Yu, Chuning He, and Yuan Jin, "Privacy-preserving technology to help millions of people: Federated prediction model for stroke prevention," [Online]. Available: Specify the source and access information.

[31] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, "Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training," CoRR, abs/1712.01887, 2017. [Online]. Available: http://arxiv.org/abs/1712.01887.

[32] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, “Gan-based synthetic medical image augmentation for increased CNN performance in liver lesion classification,” Neurocomputing, vol. 321, pp. 321–331, 2018.].

[33] M. Staffa, L. Sgaglione, G. Mazzeo, L. Coppolino, S. D’Antonio, L. Romano, E. Gelenbe, O. Stan, S. Carpov, E. Grivas et al., “An OpenNCP-based solution for secure ehealth data exchange,” Journal of Network and Computer Applications, vol. 116, pp. 65–85, 2018.

[34] D. C. Nguyen, P. N. Pathirana, M. Ding, and A. Seneviratne, “BEdgeHealth: A decentralized architecture for Edge-based IoMT networks using blockchain,” *IEEE Internet of Things Journal*, vol. 8, no. 14, pp. 11 743–11 757, 2021.