

N. RIEKE et al. The future of digital health with federated learning. *NPJ digital medicine*, v. 3, p. 119, 2020.

Caíque Santos Lima

November 3rd, 2022

Seminar MO809A



Caíque Santos Lima



PhD Candidate in Electrical Engineering

Faculdade de Engenharia Elétrica e de Computação (FEEC), 2022-2026
Computer Engineering
Viva Bem: Hub de Inteligência Artificial para Saúde e Bem Estar
Supervisor: Fernando J. Von Zuben



Master in Electrical Engineering

Departamento de Engenharia Elétrica (DEE), 2020-2022
OxiTidy: motion artifact detection-reduction in photoplethysmographic signals using artificial neural networks
Supervisor: André Carmona Hernandez
Co-supervisor: Rafael Vidal Aroca



Bachelor of Electrical Engineering

Centro Universitário Salesiano de São Paulo, 2015-2019
Campus Dom Bosco, Americana/SP

Agenda

1. Introduction
2. Objectives
3. Data-driven medicine requires federated efforts
4. Impact on stakeholders
5. Technical considerations
6. Challenges and considerations
7. Conclusion



PERSPECTIVE OPEN

The future of digital health with federated learning

Nicola Rieke^{1,2✉}, Jonny Hancox³, Wenqi Li⁴, Fausto Milletari¹, Holger R. Roth⁵, Shadi Albarqouni^{2,6}, Spyridon Bakas⁷, Mathieu N. Galtier⁸, Bennett A. Landman⁹, Klaus Maier-Hein^{10,11}, Sébastien Ourselin¹², Micah Sheller¹³, Ronald M. Summers¹⁴, Andrew Trask^{15,16,17}, Daguang Xu⁵, Maximilian Baust¹ and M. Jorge Cardoso¹²

Data-driven machine learning (ML) has emerged as a promising approach for building accurate and robust statistical models from medical data, which is collected in huge volumes by modern healthcare systems. Existing medical data is not fully exploited by ML primarily because it sits in data silos and privacy concerns restrict access to this data. However, without access to sufficient data, ML will be prevented from reaching its full potential and, ultimately, from making the transition from research to clinical practice. This paper considers key factors contributing to this issue, explores how federated learning (FL) may provide a solution for the future of digital health and highlights the challenges and considerations that need to be addressed.

npj Digital Medicine (2020)3:119; <https://doi.org/10.1038/s41746-020-00323-1>

¹NVIDIA GmbH, Munich, Germany. ²Technical University of Munich (TUM), Munich, Germany. ³NVIDIA Ltd, Reading, UK. ⁴NVIDIA Ltd, Cambridge, UK. ⁵NVIDIA Corporation, Bethesda, USA. ⁶Imperial College London, London, UK. ⁷University of Pennsylvania (UPenn), Philadelphia, PA, USA. ⁸Owkin, Paris, France. ⁹Vanderbilt University, Nashville, TN, USA. ¹⁰German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹¹Heidelberg University Hospital, Heidelberg, Germany. ¹²King's College London (KCL), London, UK. ¹³Intel Corporation, Santa Clara, CA, USA. ¹⁴Clinical Center, National Institutes of Health (NIH), Bethesda, MD, USA. ¹⁵OpenMined, Oxford, UK. ¹⁶University of Oxford, Oxford, UK. ¹⁷Centre for the Governance of AI (GovAI), Oxford, UK. ✉email: nrieke@nvidia.com

INTRODUCTION



The challenge of obtaining health data

- Health data is highly sensitive and its usage is tightly regulated
- Complexity in collecting, curating, and maintaining a high-quality dataset
- Health data (significant business value) are not freely shared
- Privacy issues¹

¹Schwarz, C. G. et al. Identification of anonymous MRI research participants with face-recognition software. *N. Engl. J. Med.* 381, 1684–1686 (2019).

AI and healthcare¹



- Drug discovery and genomics
- Disruptive innovations in radiology and pathology
- Precision medicine
- Diagnosis and treatment of disease
- Spotting malignant tumours
- Applications of NLP in healthcare
- Patient engagement and adherence applications
- Surgical robots (e.g. gynaecologic surgery, prostate surgery and head and neck surgery)

¹DAVENPORT, T.; KALAKOTA, R. The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, v. 6, n. 2, p. 94–98, 2019.

Cutting-edge applications



OBJECTIVES

Objectives

1. Discusses the key factors contributing to the transition from research to clinical practice
2. Explores how FL may provide a solution for the future of digital health
3. Highlights the challenges and considerations that need to be addressed
4. Discusses the impact on the various stakeholders in a FL ecosystem

DATA-DRIVEN MEDICINE REQUIRES FEDERATED EFFORTS

The reliance on data



- Biases in demographics (e.g., gender, age)
- Technical imbalances (e.g., acquisition protocol, equipment manufacturer)

Initiatives seeking to pool data from multiple institutions



- IBM's Merge Healthcare acquisition
- NHS Scotland's National Safe Haven
- French Health Data Hub
- Health Data Research UK

Initiatives seeking to pool data from multiple institutions



- Human Connectome
- UK Biobank
- The Cancer Imaging Archive (TCIA)
- NIH CXR828
- NIH DeepLesion
- The Cancer Genome Atlas (TCGA)
- Alzheimer's Disease Neuroimaging Initiative (ADNI)

Data Lakes challenges



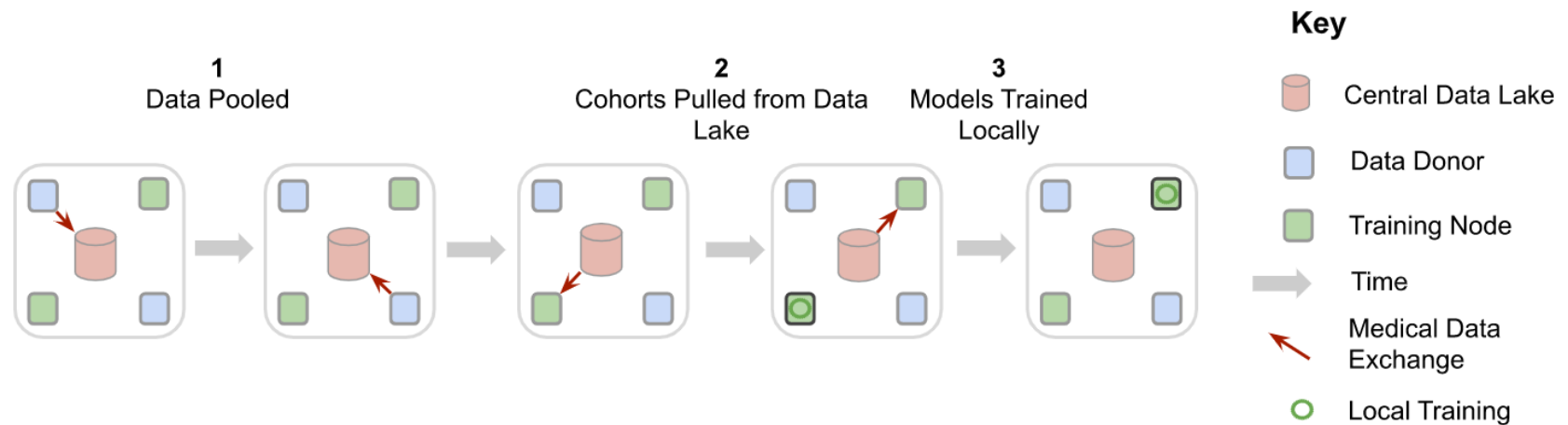
- Regulatory, ethical and legal challenges
- Privacy and data protection



The promise of federated efforts

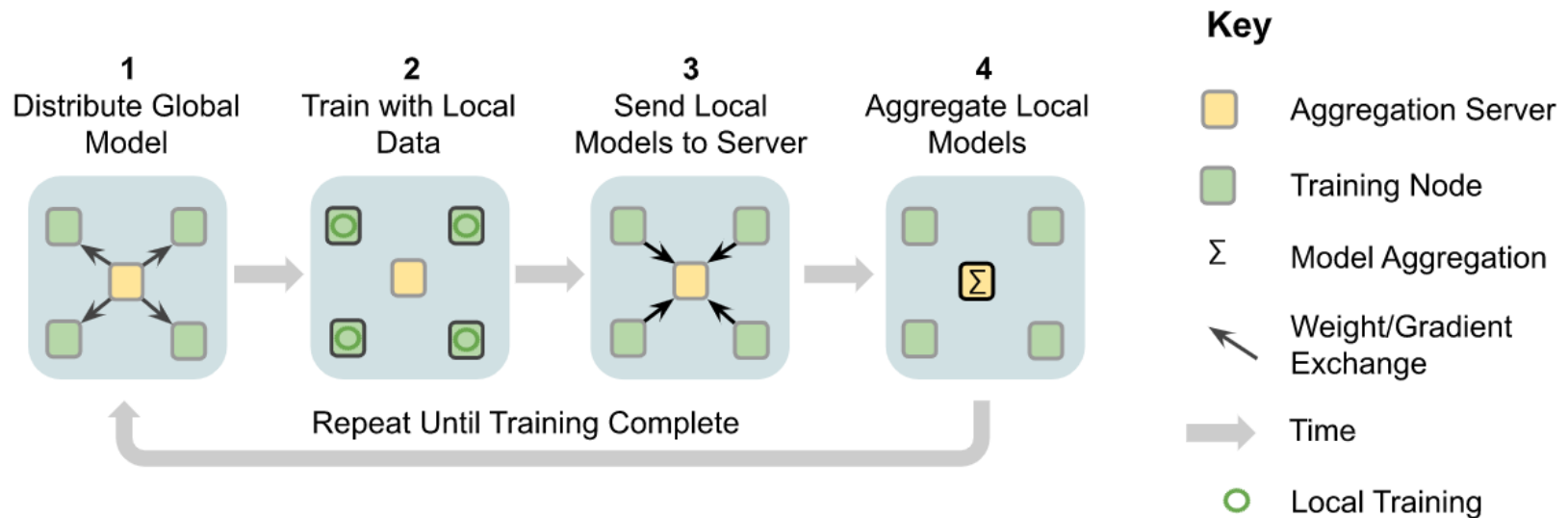
- To address **privacy** and **data governance** challenges by enabling ML from non-co-located data (moving the model to the data and not vice versa)
- Each data controller not only defines its own **governance processes** and associated **privacy policies**, but also controls data access and has the ability to revoke it

Centralised Data Lake

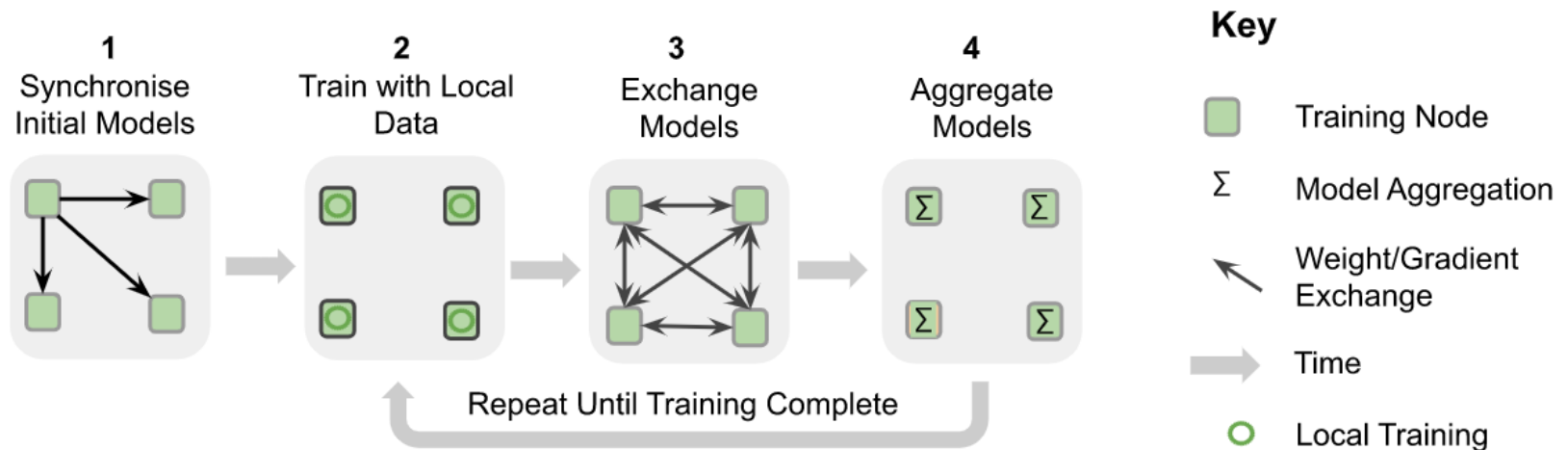




FL — Aggregation Server



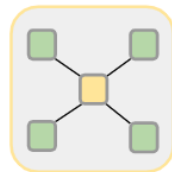
FL — Pear to Pear



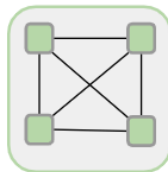
Overview of different FL design choices

FL Topologies

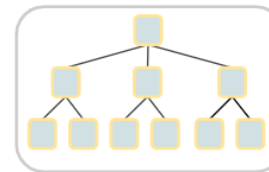
a) Centralised



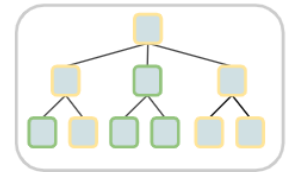
b) Decentralised









c) Hierarchical



d) Hybrid Hierarchical

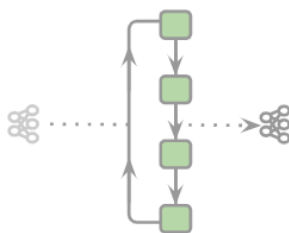


Key

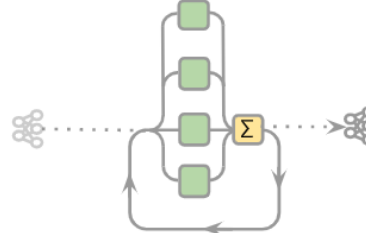
-  Aggregation Server
-  Training Node
-  Federation Nodes
- Σ Model Aggregation
-  Initial Model
-  Consensus Model
-  Weight/Gradient Exchange

FL Compute Plans

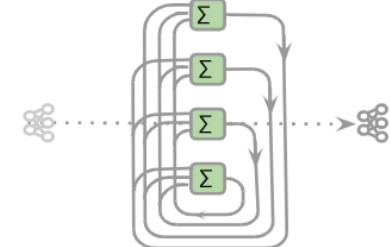
e) Sequential



f) Aggregation Server



g) Peer to Peer





Current FL efforts for digital health

- FL helps to represent and to find clinically similar patients (electronic health records – EHR)
- Predicting hospitalisations due to cardiac events
- Mortality and ICU stay time
- Whole-brain segmentation in MRI
- Brain tumour segmentation
- fMRI classification to find reliable disease-related biomarkers

IMPACT ON STAKEHOLDERS

FL ecosystem



Clinicians



Patients



Hospitals and
practices



Researchers and AI
developers



Healthcare
providers



Manufacturers

Clinicians



- Systems trained in a federated fashion are potentially able to yield even **less biased decisions** and **higher sensitivity to rare cases** as they were likely exposed to a more complete data distribution
- FL demands some up-front effort such as compliance with agreements, e.g., regarding the **data structure, annotation** and **report protocol**, which is necessary to ensure that the information is presented to collaborators in a commonly understood format



Patients



- FL on a global scale could ensure **high quality of clinical decisions** regardless of the treatment location
- Patients requiring **medical attention in remote areas could benefit from the same high-quality ML-aided diagnoses** that are available in hospitals with a large number of cases
- The same holds true for **rare or geographically uncommon diseases**, that are likely to have milder consequences if **faster and more accurate diagnoses** can be made



Hospitals and practices



- They can remain in full control and possession of their patient data with **complete traceability of data access**, limiting the risk of misuse by third parties





Researchers and AI developers

- They stand to benefit from **access to a potentially vast collection of real-world data**, which will particularly impact smaller research labs and start-ups
- Resources can be directed towards **solving clinical needs** and **associated technical problems** rather than relying on the limited supply of open datasets



Healthcare providers



- FL has the potential to **increase the accuracy and robustness of healthcare AI**, while **reducing costs** and **improving patient outcomes**, and may therefore be vital to precision medicine



Manufacturers



- Since combining the learning from many devices and applications, without revealing patient-specific information, can facilitate the continuous validation or **improvement of their ML-based systems**



TECHNICAL CONSIDERATIONS



Federated learning definition

- FL is a learning paradigm in which multiple parties train collaboratively without the need to exchange or centralise datasets
- Let \mathcal{L} denote a global loss function obtained via a weighted combination of K local losses $\{\mathcal{L}_k\}_{k=1}^K$, computed from private data X_k , which is residing at the individual involved parties and never shared among them:

$$\min_{\phi} \mathcal{L}(X; \phi) \quad \text{with} \quad \mathcal{L}(X; \phi) = \sum_{k=1}^K [w_k \mathcal{L}_k(X_k; \phi)],$$

where $w_k > 0$ denote the respective weight coefficients.

CHALLENGES AND CONSIDERATIONS



Data heterogeneity

- Medical data is particularly diverse — not only because of the **variety of modalities, dimensionality** and **characteristics in general**, but even within a specific protocol due to factors such as **acquisition differences, brand of the medical device** or **local demographics**
- Research addressing this problem includes, for example, **FedProx**¹, **part-data-sharing strategy**² and **FL with domain-adaptation**³

¹Li, T. et al. Federated optimization in heterogeneous networks. arXiv preprint arXiv:1812.06127 (2018).

²Zhao, Y. et al. Federated learning with non-iid data. arxivabs/1806.00582 (2018).

³Li, X. et al. Multi-site fmri analysis using privacy-preserving federated learning and domain adaptation: abide results. arXiv preprint arXiv:2001.05647 (2020).

Privacy and security



- Privacy vs. Performance
- Level of trust: Trusted and Non-trusted
- Information leakage

Traceability and accountability



- Traceability of all system assets including **data access history, training configurations, and hyperparameter tuning throughout the training processes** is thus mandatory

System architecture



- Ensuring data integrity when communicating by **use of redundant nodes**, designing **secure encryption methods** to prevent data leakage, or designing **appropriate node schedulers** to make best-use of the distributed computational devices and reduce idle time

CONCLUSION

Conclusion

- FL is a promising approach to obtain powerful, accurate, safe, robust and unbiased models
- As a consequence, it may open novel research and business avenues and has the potential to improve patient care globally
- FL has an impact on nearly all stakeholders
- Not all technical questions have been answered yet and FL will certainly be an active research area throughout the next decade¹

¹Kairouz, P. et al. Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*: Vol. 14: No. 1–2, pp 1-210, 06/2021.

