Applying AI to real-world health-care settings and the life sciences: Tackling data privacy, security and policy challenges with federated learning

M. Galtier, Owkin, France

D. Meadon, Owkin, United Kingdom

Introduction

Every day, millions of patient data points are collected from a wide range of sources. Harnessing the insights held within this ever-increasing volume of data in a safe, secure and ethical manner is the key to unlocking better health outcomes for everyone. Through recognising patterns at a scale unachievable by humans, machine-learning models can power the discovery of these insights. However, when it comes to health care, the journey from big data to actionable insights is fraught with challenges. Health data are sensitive and require handling with care through tight regulation. This essay explores how this challenge can be overcome through federated learning (FL), a novel machine-learning technology. FL drastically reduces privacy concerns by keeping patient data stored securely onsite during model training. The essay also discusses the salient policy implications to realise the full potential of FL technology for patients, health-care systems and life sciences companies.

The data challenges

Technology based on artificial intelligence (AI), especially machine-learning approaches, can power scientific breakthroughs at an unprecedented speed and scale. This technology can help us answer salient questions in medical research, such as: which patients should be included in clinical trials? Which molecules are the most promising targets for drug development? What is the best way to extract information from a wide variety of medical images, such as chest x-rays and scans to detect cancer or the early onset of COVID-19?

Machine-learning approaches are, however, “data-hungry”; models need access to large and diverse datasets to learn, improve accuracy and remove bias (Cahan et al., 2019). Consider a model designed to predict heart attack symptoms in the UK population. Such a model is unlikely to be widely applicable if it has only trained on the data collected at, for example, local general practices in a London suburb with a predominantly young, white population. Machine learning will not successfully transition from research

settings into everyday clinical practice without large, diverse and multimodal data (i.e. omics, digital pathology, radiology, spatial biology and clinical) (Rieke et al., 2020).

Why is it hard to gather data in health care?

Most patient data are stored at different hospitals and research centres, and distributed across servers and databases. These "silos" make it challenging for researchers and models to access enough data to train accurate and robust predictive models. Other valuable data types, such as chemical compounds used for drug discovery and development, are usually kept in-house by pharmaceutical companies. These "chemical libraries" are unlikely to be shared readily due to the competitive nature of the health-care industry.

Data scientists have to deal with heterogeneous datasets from multiple sources and in different formats. Sources include electronic health records, national or international databases, and clinical trial results. Meanwhile, formats and modalities could include medical history, laboratory test results and radiology images. This heterogeneity can make data collection and preparation for analysis tedious, lengthy and costly.

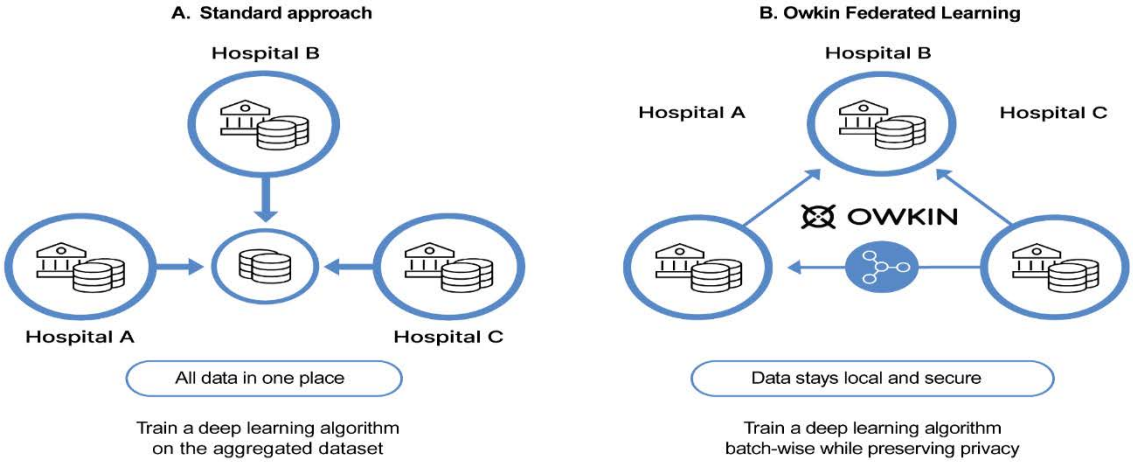
Protecting a person's identity in medical research through anonymised data is understandable but poses challenges for researchers. Completely removing identifiable information or, more commonly, using "pseudonymised" data (such as replacing a patient's name with a code) can decrease the performance of an algorithm (Rieke et al., 2020).

Consider the case of training a model to predict heart attack symptoms. This would not be clinically useful without patients' dates of birth or gender in the training dataset. Ultimately, whether anonymised or not, the data belong to patients, who may or may not consent to them being sold or transferred to third parties.

How does federated learning work?

FL emerged in 2015 to address data governance and privacy considerations by collaboratively training algorithms without exchanging the data (Figure 1). Algorithms are dispatched to different data centres, where they train locally. Once improved, they return to a central location, while the data stay local. The same algorithms are then sent to other local datasets to re-train and improve.

Figure 1. Standard machine learning versus federated learning approaches



In this way, FL enables researchers to gain insights collaboratively in the form of a consensus model without moving patient data beyond the firewalls of the institutions in which they reside. Since the machine-learning process occurs locally at each participating institution and only model characteristics are transferred, it greatly enhances patient privacy. Recent research has shown that models trained by FL can achieve performance levels comparable to those trained on centrally hosted datasets and superior to those models that only see isolated single-institution data (Huang et al., 2019).

Why is federated learning the future of health care?

The successful implementation of FL holds significant potential for enabling precision medicine at scale. Training algorithms on unprecedented quantities of heterogeneous datasets across multiple sources leads to models that yield less-biased decisions. These are sensitive to an individual's physiology while respecting governance and privacy concerns. FL still requires rigorous technical consideration for the algorithm to proceed optimally without compromising safety or patient privacy. However, it is a promising approach to creating powerful, accurate, safe, robust and unbiased models.

FL offers a solution for how to maximise the benefits of AI in health research while overcoming security and privacy issues. It is poised, therefore, to turbocharge medical research over the next few decades. In so doing, it will increase the pace and decrease the cost of developing treatments and diagnostics. At the same time, it will improve targeting and personalisation of existing treatments.

In a few years, FL will solve the problem of siloed datasets. It will allow diverse datasets from across the health sector, including from health-care settings and clinical trials, to be rapidly, safely and securely analysed. Instead of developing machine learning from scratch, foundation models for each medical field will emerge, enabling an ontology of tools and models to solve more specific problems. The impact of AlphaFold on the understanding of the protein structure will be replicated across different areas – from treatments to proteins to virus genomes.

From a patient's perspective, these FL-enabled foundation models will lead to clinical models that can be systematically checked by the US Food and Drug Administration (FDA) and other agencies. It will self-update, continually integrating and verifying clinical models that can be implemented in any medical device or clinical trial.

As patients become increasingly empowered to control their data, regulators are also implementing more rigorous data and privacy controls. In the coming decades, a collaborative process will likely be needed to publish clinical machine-learning models. The FDA would screen and verify datasets, and systematically remove any biases. Researchers will benefit from a decentralised but single source of truth for model building using FL.

What is happening in the federated learning space?

Recent years have seen a surge in real-world applications of FL in health care. The COVID-19 pandemic required scientific, academic and medical collaboration at an unprecedented scale. This accelerated the data sharing necessary to expedite critical research with direct impact on patient care. In one study, scientists leveraged the NVIDIA Clara FL platform to train the EXAM (electronic medical record chest X-ray) FL model across heterogenous, unharmonised datasets from 20 institutes. This allowed them to predict the future oxygen requirements of symptomatic patients with COVID-19. This, in turn, can aid physicians in determining the appropriate level of care (Dayan et al., 2021). NVIDIA Clara has also been used for a broad range of AI applications in medical imaging, genetic analysis and oncology. This includes segmenting pancreatic tumours or training a mammography classification model across five institutions (Wang et al., 2020).

New technology is emerging as the global research community looks to unlock the potential of FL. Large corporates including IBM have launched their own FL frameworks to build adverse drug reaction prediction models using electronic health data, among other applications (Choudhury et al., 2019). A proliferation of start-ups in the space offer their own potential platforms and solutions, but few have managed to apply these in real-world settings at scale.

The public sector has also become increasingly active. The UK Government, for example, outlined a plan to set up a federated infrastructure for management of UK genomics data. This would bring them together with other routinely collected health data to support work on cancer and rare diseases (UK Government, 2021).

Owkin Connect is an FL software that powers collaborations between hospitals, research centres, technology partners and life science companies (Owkin, 2022). It is used for clinical research in the HealthChain consortium and for drug discovery in the MELLODDY consortium (MELLODDY, 2022). The MELLODDY project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking. This, in turn, receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations. The HealthChain project has received funding from the French public investment bank (Banque Publique d'Investissement).

Life sciences

Enabled by an ever-expanding arsenal of model systems, analysis methods, libraries of chemical compounds and other agents (like biologics), the amount of data generated during drug discovery programmes has never been greater. However, the biological complexity of many diseases still defies pharmaceutical treatment. Coupled with the rise in regulatory expectations, this growing complexity has inflated the research intensity and associated cost of the average discovery project. It is, therefore, imperative to maximise learning from these data investments.

Owkin is the project co-ordinator for MELLODDY, a consortium of ten pharmaceutical and seven technical partners funded by the Innovative Medicines Initiative (IMI, 2022). The partners include well-known firms such as Janssen, AstraZeneca, Novartis, GSK, Bayer, Amgen, Astellas, Kubermatic, NVIDIA and KU Leuven. This landmark project demonstrated that collaborating in AI for drug discovery is possible at industrial scale.

Launched in 2019, MELLODDY announced the successful development and operation of a secure platform for FL without sharing each partner's proprietary data and models or compromising their security and confidentiality in 2020. In 2021, MELLODDY shared the first-ever demonstration of the predictive performance benefits of FL in drug discovery. The final results published in 2022 show that its operation at scale yields improvements across all pharmaceutical partners in the predictive performance of collaboratively trained models over single partner models. Models that more accurately predict the pharmacological and toxicological activities of molecules better support the decision-making process of which candidate drug molecules to make and test.

Clinical research

Through collaborative projects like HealthChain, Owkin has successfully trained machine-learning models on histology images, siloed at different clinical centres, to predict treatment responses in breast cancer. A model trained with Owkin Connect can now help oncologists select the most effective breast cancer treatment for each patient based on a single biopsy and identify high-risk patients for clinical drug trials.

Federated learning and data privacy

Ensuring the privacy and security of personal data has become a significant challenge posed by the rise of AI and machine learning. The standard approach of centralising data from multiple centres requires datasets to reside on a single server. This often means uploading private data to the cloud, sharing data with other parties and transferring vast amounts of data to a data centre for processing. Because FL is decentralised by design, and privacy preserving, it enables compliance with the General Data Protection Regulation (GDPR). The table below highlights the key difference in the impacts on data privacy between centralised machine learning and FL.

Table 1. Centralised machine learning versus federated learning impacts on data privacy

Centralised machine learning	Federated learning
With data merging, a personal data violation by a third-party intrusion can impact the whole dataset and paralyse research until the violation terminates.	Federated learning prevents suspension of the research, paralysing research only on the violated site. The violation concerns a limited set of data only.
The ability to centralise data in one dataset can impact personal data processing as it requires systematic production of a data impact assessment.	Data stay onsite, and interconnection is not systematic. In this situation, sites do not always require a data impact assessment.
Sites are identified as data co-controllers: a co-controller contract needs to be drafted and negotiated between the different sites.	The different sites are not identified as data co-controllers as the data processing is done between the site and the data processor (Owkin). As such, data processing is facilitated and there is no need to draft a co-controller contract.

Note: Machine-learning model parameters exchanged between parties in an FL system still conceal sensitive information, which privacy attacks can exploit. It is therefore crucial that federated learning is bolstered with efforts to ensure gold-standard privacy preservation. Owkin's software stack includes secure aggregation. This distributed cryptographic method, a combination of multi-party computation and homomorphic encryption (a form of encryption that permits users to perform computations on its encrypted data without first decrypting it), enables encrypted models to be averaged without being decrypted. As a consequence, the models are far more protected because only aggregated models are shared across network partners and the platform operator (Owkin) cannot decrypt the models.

Solving policy challenges using federated learning

FL can square the circle of harnessing the value of health-care data, while preserving patient privacy. This improved data security is critical in Europe. The GDPR has put individual privacy at the heart of the continent's approach to data governance. Moreover, the European Union will further strengthen its data protection rules through ePrivacy Regulation. This regulation will introduce strict obligations around data confidentiality and require user consent before metadata can be processed.

However, this is not just about Europe. Across the world, jurisdictions are following the European Union's lead in introducing sweeping new privacy frameworks. In 2022, the People's Republic of China (hereafter "China") enacted the Personal Information Protection Law. This comprehensive GDPR-like regulation imposes restrictions on data collection, transfer and analysis (Xu et al., 2021). Companies inside and outside of China will need to comply. In the United States, numerous individual states have passed their own privacy laws. Meanwhile, the introduction of federal regulation is increasingly a matter of "when" rather than "if". In other words, data innovation and respect for personal privacy will increasingly go hand in hand wherever one operates. By adopting FL, organisations can put themselves in a solid position to comply with present and future privacy regulations.

FL also aligns with a growing priority among policy makers: ensuring that data are stored and processed locally – otherwise known as data localisation or "data sovereignty". While data localisation requirements are more common in partner economies such as China and the Russian Federation, data localisation demands are becoming increasingly prominent among European and US policy makers (Svantesson, 2020).

The motivations behind these demands vary. Some want to safeguard national security. Others lack trust in the data protection standards of jurisdictions other than their own. Still others believe that storing and processing data locally is – or will soon be – required for economic competitiveness.

FL responds to these issues. It enables data to be analysed by algorithms where they are stored – rather than requiring them to be transferred and pooled elsewhere. It thus solves "poli-analysed" concerns, while safeguarding innovation.

FL also has a crucial role in helping policy makers unlock the full economic and social value of data. The experience of the COVID-19 pandemic has illustrated, like never before, the ability of data to help humanity solve complex collective challenges. These could range from tracking the emergence and spread of new variants to monitoring vaccine side effects in billions of individuals.

To bolster these capabilities, governments can take steps to harness the power of data across various areas – from treating disease and reducing emissions to combating financial fraud. For example, in May 2022, the European Commission presented its much-anticipated Health Data Space (HDS) (European Health Data Space, 2022). This health-specific ecosystem is comprised of rules, common standards and practices, infrastructures and a governance framework. It aims to empower individuals through increased digital access to, and control over, their electronic personal health data. At the same time, it fosters a genuine single market for electronic health records systems, relevant medical devices and high-risk AI systems.

The HDS also aims to provide a consistent, trustworthy and efficient set-up for the use of health data for research, innovation, policy making and regulatory activities (secondary use of data). The intention is to promote greater data sharing between businesses, researchers and public institutions to drive innovation and economic growth. FL provides the mechanism for achieving these aspirations. To that end, it enables insights generated from multiple datasets, while avoiding the technical challenges, confidentiality concerns and privacy risks that inevitably arise when pooling datasets.

Perhaps most important of all, through its emphasis on data privacy and security, FL could contribute to rebuilding the general public's damaged trust in the societal benefits of technology. This trust eroded in recent years due to the widespread sense that technology companies have abused the access to data that consumers have granted them. They believe these firms either failed to protect the confidentiality of that data or monetised them without meaningful user consent. The widespread adoption of FL could address these legitimate concerns and increase the public's willingness to share data that can drive innovation in health care and many other areas.

Conclusion

By enabling multiple parties to train collaboratively without the need to exchange or centralise datasets, FL addresses issues related to the transfer of sensitive medical data. Consequently, it opens novel research and business avenues and can improve patient care globally. FL is already having a significant impact on the health-care ecosystem. It is offering better diagnostic tools to clinicians by improving the analysis of medical images and other clinical data. It is driving precision medicine by helping identify patient subgroups to accelerate clinical trials. This acceleration decreases the cost and time-to-market for pharmaceutical companies, ensuring more patients receive the right treatment faster.

As a relatively new technology, FL will undoubtedly be an active research area throughout the next decade. It already offers exciting new opportunities for research breakthroughs by overcoming fundamental privacy concerns. With valuable use cases in health care expected to continue to emerge, more health-care partners will decide to use FL to collaborate safely and securely. This will lead to a true paradigm shift to make precision medicine a reality and ultimately improve patient care. By safely and ethically unlocking

the invaluable insights stored within patient data, FL will increasingly play a crucial role in ensuring every patient gets the right treatment.

However, this scenario may require some appetite from the public sphere. In geographies with stringent approaches to patient data usage, FL might be critical in facilitating access to data for secondary use, as well as for sensitive data like genomics. As an alternative to the dominant "hub approach" in which data are aggregated in a single place, FL's decentralised approach and cryptographic features offer political decision makers and public authorities a compelling solution to cybersecurity issues.

This public appetite may first come through public financing, especially in helping research centres to adopt a decentralised approach and to create shared infrastructure. It may also come through regulation, especially at the EU level. In that perspective, the EU HDS constitutes a strong opportunity to further increase policy makers' awareness on FL and to adopt a decentralised strategy.

References

Cahan, E.M. et al. (2019), "Putting the data before the algorithm in big data addressing personalized healthcare", *npj Digital Medicine*, Vol. 2/78, https://doi.org/10.1038/s41746-019-0157-2.

Choudhury, O. et al. (2019), "Predicting adverse drug reactions on distributed health data using federated learning", presentation to AMIA Symposium, https://research.ibm.com/publications/predicting-adverse-drug-reactions-on-distributed-health-data-using-federated-learning.

Dayan, I. et al. (2021), "Federated learning for predicting clinical outcomes in patients with COVID-19", *Nature Medicine*, Vol. 27, pp. 1735-1743, https://doi.org/10.1038/s41591-021-01506-3.

EC (2022), "European Health Data Space", webpage, https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space_en (accessed 25 November 2022).

Huang, L. et al. (2019), "Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records", *Journal of Biomedical Informatics*, Vol. 99/103291, https://doi.org/10.1016/j.jbi.2019.103291.

IMI (2022), Innovative Medicines Initiative website, www.imi.europa.eu/ (accessed 25 November 2022).

MELLODDY (2022), MELLODDY website, www.melloddy.eu/ (accessed 25 November 2022).

Owkin (2022), "Owkin Connect", webpage, www.owkin.com/connect (accessed 25 November 2022).

Rieke, N. et al. (2020), "The future of digital health with federated learning", *npj Digital Medicine,* Vol. 3/119, https://doi.org/10.1038/s41746-020-00323-1.

Svantesson, D. (2020), "Data localisation trends and challenges: Considerations for the review of the Privacy Guidelines", OECD Digital Economy Papers, No. 301, OECD Publishing, Paris, https://doi.org/10.1787/7fbaed62-en.

UK Government (2021), "Genome UK: 2021 to 2022 Implementation Plan", www.gov.uk/government/publications/genome-uk-2021-to-2022-implementation-plan/genome-uk-2021-to-2022-implementation-plan.

Wang, P. et al. (2020), "Automated pancreas segmentation using multi-institutional collaborative deep learning", *arXiv,* arXiv:2009.13148 [eess.IV], https://arxiv.org/abs/2009.13148.

Xu, K. et al. (2021), "Analysing China's PIPL and how it compares to the EU's GDPR", International Association of Privacy Professionals", 24 August, https://iapp.org/news/a/analyzing-chinas-pipl-and-how-it-compares-to-the-eus-gdpr.