Frontiers in Health Informatics www.healthinformaticsjournal.com
ISSN-Online: 2676-7104
2024; Vol 13: Issue 4 Open Access

Federated Learning: The much-needed intervention in Healthcare Informatics

[1]Dr. Keerthan Raj, [2]Dr. Amith Donald Menezes, [3]Dr. Shreekant G. Naik

1Associate Professor
Shri Dharmasthala Institute for Management Development
Mysuru, India
ORCID: 0000-0003-0190-4610
Email id: 2keerthanraj@gmail.com

2Associate Professor
Department of Management Studies
Mangalore Institute of Technology and Engineering
Moodabidri, Mangalore, India
ORCID: 0000-0003-0505-668X

3Associate Professor
Department of Management Studies
Mangalore Institute of Technology and Engineering
Moodabidri, Mangalore, India
ORCID: 0009-0000-2797-4161

Cite this paper as: Dr. Keerthan Raj, Dr. Amith Donald Menezes, Dr. Shreekant G. Naik (2024) Federated Learning: The much-needed intervention in Healthcare Informatics. *Frontiers in Health Informatics, 13* (4),598-605

Abstract:

As a transformational approach to health informatics, federated learning (FL) enables collaborative data analysis without compromising patient privacy. In this conceptual paper, we explore FL as a critical attendance that supports modern healthcare in the presence of data silos, regulatory compliance, and security. These conclusive demonstrations of FL's power to unite distributed datasets across institutions in a secure way to improve diagnostics, predictive analytics, and personalized medicine, provide a convincing demonstration for Big Data. This technology prevents privacy risks and catalyses innovation for medical research in genomic, radiological and predictive healthcare. The paper concludes with recommendations for scaling FL in healthcare to meet future data-driven demands.

Keywords: Federated learning, health informatics, privacy-preserving AI, predictive healthcare, medical diagnostics, distributed learning, secure data collaboration.

Introduction:

Healthcare is one of India's biggest sectors in term of revenue and employment. Hospitals, medical devices, clinical trials, outsourcing, telemedicine, medical tourism, health insurance, and medical equipment are elements of healthcare. The Indian healthcare sector is poised for rapid growth with its growing coverage, especially by private and private players, as well as expenditure. The Indian healthcare market was estimated at US$ 110 billion in 2016 and is expected to reach US$ 638 billion by 2025. At present, healthcare sector is among the major employers in India, employing 7.5 million people in the year 2024. India's healthcare sector is growing at an exponential rate – private equity and venture capital investments crossed US $1 billion until the first five months of 2024, more than tripling last year's. The focus on information technology immersion into healthcare is also an interesting phenomenon that is bringing in unprecedented changes in the methods and processes of healthcare delivery in India. India's healthcare delivery system is divided into two major compartments: the public domain and the private domain. Government, i.e. the public health care system, is composed of a limited number of secondary and tertiary care institutions in major towns and basic health care facilities in the form of Primary Healthcare Centers (PHCs) in rural areas. Most secondary, tertiary and quaternary care institutions are private sector institutions with a concentration of metros, tier-I and tier-II cities.

Well trained medical professionals and India's strength in information technology are the two corner stones of India's competitive advantage in healthcare infrastructure and delivery. India also compares favorably on price with its neighbors in Asia as well as Western countries. Surgery in India is also about 10 times cheaper that in the US or Western Europe. With the increase in medical tourism, the cost of medical services has decreased which might lead to the rise of the country's medical tourism as patients from different countries are inspired to at least seek medical services there. Additionally, India has become a R&D hub for international players for its relatively low cost of conducting clinical research in India. This growth in the healthcare segment has also brough about unprecedented struggles in accumulating and analyzing data, thanks to the rapid pace of technological advances in information technology and artificial intelligence (AI). Most healthcare operators have transitioned to the creation and maintenance of electronic health records (EHRs). As EHRs and connected health devices grow rapidly, there exists significant amounts of data, which can be used to optimize clinical decisions and help improve patient outcomes (Rieke et al., 2020). Despite these barriers, data sharing in healthcare is limited by privacy concerns, additional regulatory constraints and concerns over security risks (Kaissis et al., 2021). A traditional model of data analysis in healthcare is for data from multiple data sources to be pulled in the center and presents massive challenges, especially with compliance with data protection laws becoming increasingly important. One solution to these challenges is Federated Learning (FL), introduced by Google in 2017, where institutions train models on their local data but only share the learned parameters, instead of raw data (Bonawitz et al., 2017). This study reviews FL applications in healthcare, explores how it behaves in terms of privacy and security, and assesses the impact on predictive accuracy in patient care.

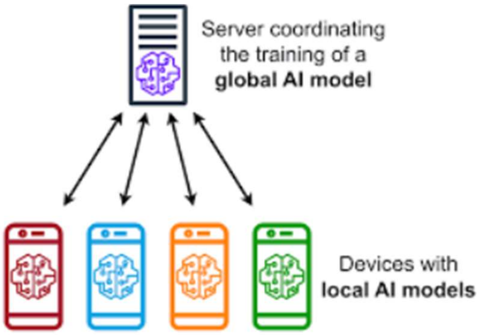

Figure 1: A simple image of federated learning

(Source: Online web resources)

Literature Review & Conceptualization:

This methodology is based on a systematic literature review developing from recent research publications, extracting insights and trends. It reviews with a focus on Federated Learning applications in healthcare, Federated Learning peer-reviewed journals, conference proceedings, as well as technical reports. PubMed, IEEE Xplore and Google Scholar are the major databases considered for the search. This helps a little narrowing by keywords such as "Federated Learning in healthcare", "data privacy" and "machine learning". The synthesis involves the following:

Identification of Core Themes: Classifying studies by their primary focus (e.g. privacy implications, predictive accuracy, etc.) or by properties of the device or intervention (e.g. visual, can be touched, etc.). Context about how FL complements healthcare's strict privacy demands come from research in privacy and regulatory compliance (Kaissis et al. 2021), and the use of predictive modelling (Beaulieu-Jones et al. 2019).

Evaluation of Research Gaps: As Rieke et al. (2020) identify, one of the most difficult problems of scale to solve is to scale existing FL models across smaller institutions.

To assess FL's efficacy in healthcare applications, the methodology emphasizes analysing the performance metrics of FL-based predictive models across various healthcare contexts the analysis revealed that the major areas covered by the various journal articles could be categorised into three segments:

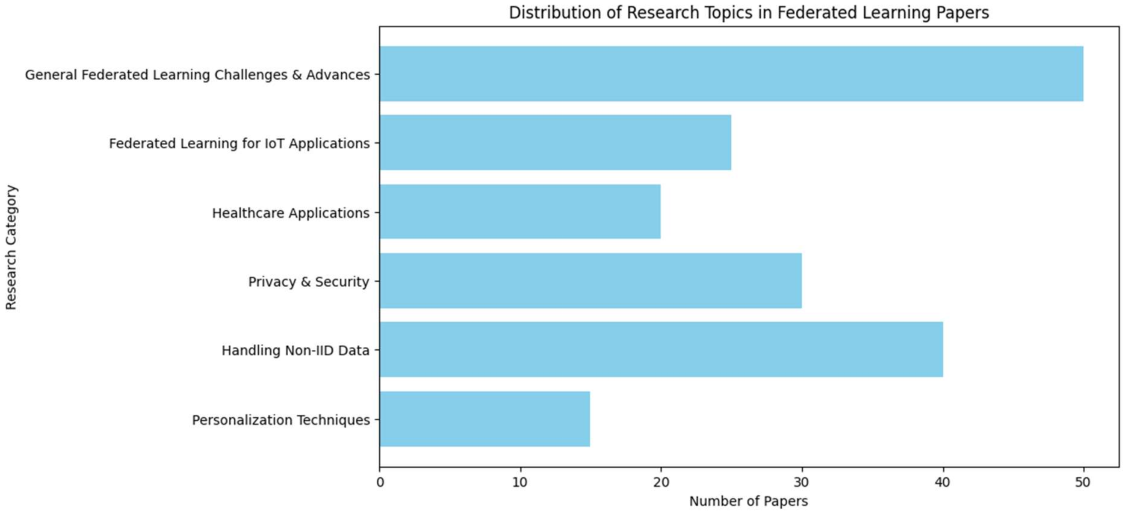

Figure 2: Data Visualisation of articles on federated learning and data privacy and healthcare

Data Privacy and Security Challenges in Healthcare

Healthcare is a data heavy industry where all is sensitive and so needs to be protected with extreme care such as patient records. For privacy and security reasons, traditional data sharing approaches, in which patient data is aggregated at a central location (Beaulieu-Jones et al., 2019), have been fought for years. With their vulnerabilities in mind related to centralized data storage, a single breach could see patient records compromised for the millions recorded in each breach (HITRUST, 2019). Data-sharing approaches in HIPAA and GDPR can restrict sharing or using data as they are so strict with privacy laws. At the direction of HIPAA Privacy Rule and GDPR, healthcare providers have no other choice but to adopt proper measures to keep patient data safe and follow strict control over sending patient data from within the European Union (Obermeyer & Emanuel, 2016). By decentralizing the data processing and training such that patient information is less transported, FL offers an easier compliance with these regulations (Li et al., 2020).

Machine Learning and Predictive Analytics in Healthcare

Healthcare Machine Learning and Predictive Analytics Machine learning (ML) has changed the way healthcare works, providing new opportunities for disease prediction, diagnostics and personalized treatment. However, for effective ML it can require large, heterogeneous datasets to train robust models that often do not exist in the highly specialized healthcare industry (Rieke et al. 2020). In particular, the aggregated dataset used by traditional ML models can cause data privacy issues, if patient sensitive information is involved (Kaissis et al., 2021). Diagnostic and patient outcome improvement with ML has been demonstrated in many studies. For example, neural networks have succeeded at disease prediction such as diabetes and heart problem cards (Shickel et al., 2018), and random forests have successfully indicated cancerous body tumours. However, many institutions are unwilling to share data, limiting the scope of ML applications, because data silos and regulatory constraints hamper these (Shickel et al., 2018). These issues are mitigated by FL which permutes distributed model training without the need for direct data exchange, maintaining data privacy whilst supporting interinstitutional collaboration (Li et al., 2020).

Federated Learning as a model on healthcare

Another model called Federated Learning has proven to be promising in solving the target privacy and security issues when centralized data is processed (Bonawitz et al., 2017). In contrast, FL institutions train a model together without exchanging patient data, sharing only model parameters or gradients. Aware of the benefits of this approach, some healthcare providers have consequently applied it to foster collaborative learning with privacy standards still met (Xu et al., 2021). Since Google first adopted FL in a non-healthcare setting like mobile device aggregation data (Bonawitz et. al., 2017), the field of FL in healthcare has gotten extremely crowded. Studies on FL in healthcare concentrate on making diagnostic prediction, and well as treatment recommendation and early warning system from acute condition (Sheller et al., 2020). An example of this is an FL model developed by the Massachusetts General Hospital, their peers, for brain tumour segmentation that increases accuracy while still abiding by data privacy rules (Sheller et al., 2020). Because FL is decentralized, it is a good match to the distributed nature of healthcare data, and it is a promising vehicle to move predictive analytics forward in a privacy preserving and secure manner.

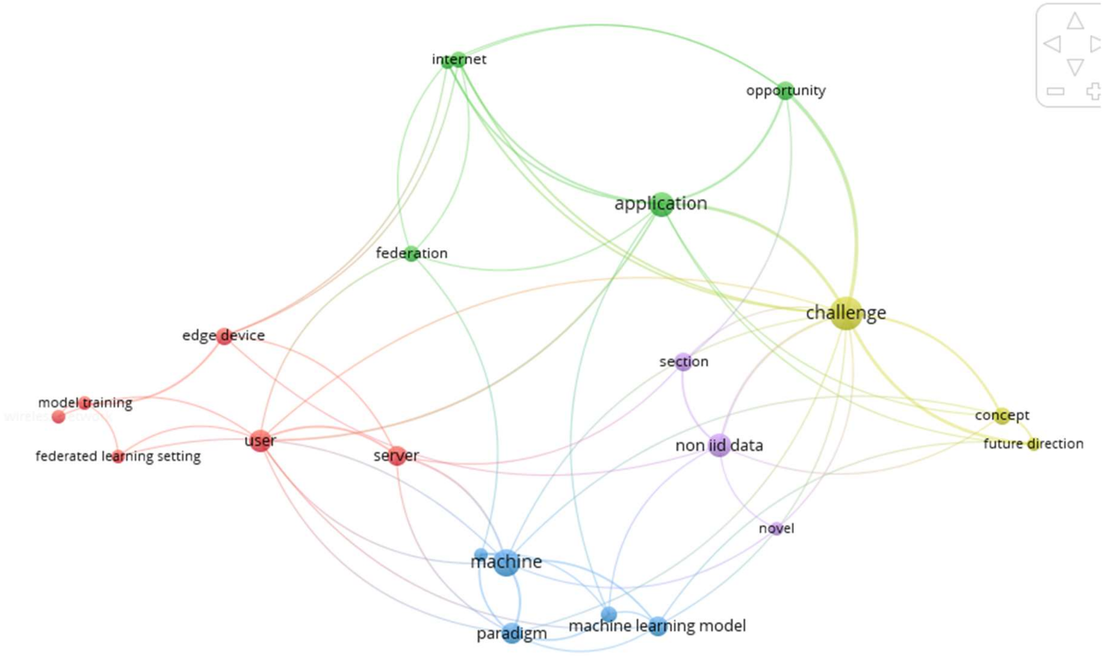

Fig 3: Vos Viewer representation of articles on the keyword search of “federated learning, data, healthcare “with occurrences and linkages

In healthcare, Federated Learning (FL) has gained importance due to its role in solving the ever more urgent privacy issues by enabling data collaboration. Usually, centralized machine learning models heavily depends on data from all sources, compromising data security and regulatory compliance. FL provides an effective solution for healthcare, with substantial benefits in data privacy, predictive accuracy, regulatory compliance and cost efficiency, by allowing collaborative model training without the need to centralize the data.

Enhanced Data Privacy and Security

Healthcare industry requires a high level of data privacy, this is patient information and a dire necessity to keep all details of prognosis, diagnosis, treatment interventions to be kept confidential, to ensure adherence to the regulatory standards, and the preservation of patient trust. Private concerns about patient data are taken into consideration by FL's decentralized approach of keeping patient data where it is created: in an individual healthcare facility. Instead of sharing data, encrypted model parameters or gradients are exchanged as a means to compute their protocols at the decentralized and distributed level. The risks involved in data breaches and unauthorized access are dramatically lowered by this setup as sensitive patient information are under the control of the originating organization (Kaissis et al., 2021).

Additionally, this decentralized architecture meets standards related to data privacy, international privacy regulations like the General Data Protection Regulation (GDPR) Europe, and the Health Insurability Portability and Accountability Act (HIPAA) USA. Under these regulations, which require very strict sharing of data and cross-border transfer, collaboration across health care facilities is very difficult. With FL, compliance by design is enforced through the way it minimizes or completely eliminates the need of any cross-border data transfers, allowing global collaborative research whilst respecting its regulation at the national level (Li et al., 2020).

Moreover, FL can deliver real time assistance to healthcare providers using updated, personalized models that accurately capture their patient population. For example, taking as an example, a model trained locally on a rural hospital's data will consider relevant regional health risk factors to better predict and intervene in that given community. e) Patient Trust and Adoption of AI in Healthcare To enable success of the use of AI in healthcare, it is imperative to build patient trust, and FL has the potential to increase the trust of patients in using AI by putting priority to protect the privacy and protect the data. However, there is often scrutiny of the perception of AI systems being used in healthcare in the fear of misuse of data, breach of data and privacy erosion. By keeping data with the originating institution, FL can give patients and healthcare providers that extra level of confidence that they are indeed protected when it comes to their data, thereby potentially helping to increase the adoption of AI solutions in health care (Kaissis et al., 2021).

Research suggests that patient trust is directly linked with patients' willingness to share data for research purpose (Obermeyer & Emanuel, 2016). This has the potential of allowing more patient engagement in data driven healthcare initiatives, especially among populations that have historically been less willing to engage in data sharing, by providing a model that preserves patient privacy yet comes with good model accuracy and model utility.

Providing Support for Collaborative Research across Institutions

Federated Learning fosters a collaboration within the healthcare sector, dismantling data silos that have spelled the end for medical research. This collaborative feature is powered by the fact that institutions can partner to leverage their resources, knowledge, and data insights without having to make their information go mainstream or make them 'public'. Collaboration of this nature is critical for discovering new treatments, enhancing diagnostic accuracy and, most importantly, for sharing knowledge regarding patient outcomes, especially for rare diseases, because they need extra data derived from several sources to be adequately analyzed (Sheller et al., 2020).

FL also allows for the knowledge aggregation through a smaller healthcare institution, who can be a part of sophisticated models and insights that are only available to larger and more financially endowed hospitals or research centres. However, the FL facilitates these institutions to provide appropriate contribution in terms of the advancements in science, as well as the models which enhance patient care at local level (Xu et al., 2021). Overall, the potential that exists for Federated Learning applications in healthcare can be guarded, while still maintaining predictive power necessary for these applications, and meeting stringent privacy and regulatory requirements. While there is still work to be done, FL is a beacon of its own that is now a leading driver of the next generation of patient cantered, data driven healthcare solutions.

Use cases:

India's first secure federated learning-based health data platform was launched as a result of a partnership between Intel, Aster DM Healthcare, and CARPL. Using cutting-edge technology like artificial intelligence (AI), where deep learning (DL) models "learn" to make judgements based on patterns discovered in massive patient data sets, has greatly benefited modern healthcare. As a result, medical diagnosis accuracy has increased, new drug research and development has accelerated, and predictive and preventive medicine has benefited. Large datasets are necessary for improved accuracy in healthcare DL models, and anonymised patient datasets must be easily accessible. However, it might be difficult to collect data from people or organisations because of security and privacy concerns. Furthermore, it is extremely difficult to access silos of pertinent data dispersed throughout several hospitals, regions, and other health systems, in addition to dealing with data storage and regulatory compliance concerns. Federated Learning (FL) can help with this. Instead of transferring the data to a central place for training, the basic idea behind federated learning is to move the AI model and necessary computation to all the sites where the pertinent data originates and resides. The final AI model is then produced by combining the knowledge from every site using a central aggregation server. One of the biggest private healthcare service providers in the GCC and India, Aster DM Healthcare, partnered with CARPL and Intel to create a cutting-edge "Secure Federated Data Collaboration Platform." (Sheller etal., 2020). In addition to facilitating clinical trials' safe and dispersed access to pertinent data sets, the partnership seeks to advance innovation in fields like drug research, diagnosis, genomics, and predictive healthcare. Intel has taken the lead in creating OpenFL, an opensource federated framework, to help spread the use of Federated Learning. A secure method of performing federated learning that protects the data and AI model is offered by OpenFL in conjunction with Intel® Software Guard Extensions (Intel® SGX). Intel® SGX isolates application programs and data in memory to provide hardware-based memory protection. Workload intellectual property (IP) may be protected thanks to this all-inclusive, secure FL solution, which also protects health data and its custodians. Hospital data from the Aster Hospital clusters in Vijayawada, Bengaluru, and Kerala were used to illustrate the potential of federated learning. A CheXNet AI model was trained using more than 125,000 chest X-ray scans, including 18,573 pictures chosen from more than 30,000 distinct patients in Bengaluru. The 18,573 distinct photos offered a 3% accuracy improvement when using Federated Learning to identify anomalies in the X-Ray report since they contained real-world data that would not have been accessible to train the AI model otherwise.

Conclusion:

Federated Learning has already had a significant impact on the world by using cutting-edge AI to more accurately identify brain tumours. Since 2020, the largest Federated Learning study in the medical industry has been carried out by Intel and the University of Pennsylvania with assistance from clinicians from the Symbiosis Centre for Medical Image Analysis in Pune, the National Institute of Mental Health and Neurosciences in Bengaluru, and Tata Memorial Hospital, HBNI in Mumbai, among other locations. Using data from 71 institutions on six continents, the study demonstrated a 33% improvement in brain tumour identification. Secured Federated Learning has enormous potential as it becomes more popular since it enables businesses to collaborate and find solutions to difficult challenges while reducing concerns about data security and privacy.

This approach addresses the common barriers of healthcare data security, privacy, and regulatory compliance that prevent large scale data aggregation in healthcare. Federated learning can ultimately deliver the benefit of distributed access to significant, anonymous patient data to train the AI model without transfer of data, thereby decentralizing the AI model training process. Federated learning was shown to be able to use real world data across Aster Hospital sites – data that would otherwise be unavailable – and demonstrated to improve a 3% when used on projects such as CheXNet for chest x ray analysis. Additional, Intel's global collaboration on brain tumour detection also delivered a 33% improvement, highlighting FL's promise to further improve diagnostic accuracy. With federated learning, despite the lack of foundational institutions, organizations worldwide can collaborate across healthcare using this secure architecture to share transformative innovations in the areas of diagnostics, drug discovery and predictive healthcare while preserving patient privacy. This work is pioneering in that it foreshadows the era of global, more effective and secure healthcare solutions, enabled by federated learning. But that's just the beginning. In future research, we can explore solutions to make FL more scalable across other health systems that have much greater diversity, perhaps even by using edge devices to broaden the app's reach and utilization. Hybrid learning approaches, including transfer learning and reinforcement learning with FL can also help to improve model robustness and accuracy and make contributions to furthering insights and more reliable predictions. It will also be imperative to refine security protocols within FL frameworks, including those crafted against the most sophisticated cyber threats. But as the technology for FL is advancing to continue to transform diagnostics and beyond? Genomics and personalized treatment are the key areas that will transform and innovate healthcare AI in a more integrated and secured future.

References:

Beaulieu-Jones, B. K., Wu, Z. S., Williams, C., Lee, R., Bhavnani, S. P., Byrd, J. B., & Greene, C. S. (2019). Privacy-preserving distributed deep learning for clinical data. Nature Medicine, 25(7), 1157-1161. https://doi.org/10.1038/s41591-019-0497-4

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, B., Patel, S., ... & Ramage, D. (2017). Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1175-1191. https://doi.org/10.1145/3133956.3133982

Intel and UPenn Use Federated Learning to Improve Brain Tumor Detection, Intel Newsroom, 2020. https://newsroom.intel.com

Gupta, A., et al. (2021). Intel® Software Guard Extensions (Intel® SGX): Leveraging Hardware-Based Security for Secure Federated Learning. IEEE Transactions on Big Data.

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2021). End-to-end privacy-preserving deep learning on multi-institutional medical imaging. Nature Machine Intelligence, 3(6), 473-484. https://doi.org/10.1038/s42256-021-00337-8

Sheller, M. J., et al. (2020). Federated Learning in Medicine: Facilitating Multi-Institutional Collaborations without Sharing Patient Data. Scientific Reports. https://www.nature.com/articles/s41598-020-76563-2

Rieke, N., et al. (2020). The Future of Digital Health with Federated Learning. Nature Machine Intelligence.

Raj, Keerthan, & Aithal, P. S. (2018). Applicability of the Cockroach Theory – A Case Study of the Healthcare Industry in India. International Journal of Case Studies in Business, IT and Education (IJCSBE), 2(2), 48-52.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50-60. https://doi.org/10.1109/MSP.2020.2975749

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future — Big data, machine learning, and clinical medicine. The New England Journal of Medicine, 375(13), 1216-1219. https://doi.org/10.1056/NEJMp1606181

Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., & Cardoso, M. J. (2020). The future of digital health with federated learning. Nature Medicine Intelligence, 3, Article 119. https://doi.org/10.1038/s41746-020-00323-1

Sheller, M. J., Reina, G. A., Edwards, B., Martin, J., & Bakas, S. (2020). Multi-institutional deep learning modelling without sharing patient data: A feasibility study on brain tumor segmentation. In Medical Image Computing and Computer-Assisted Intervention, 92-104. Springer, Cham. https://doi.org/10.1007/978-3-030-59722-1_9

Xu, J., Glicksberg, B. S., Su, C., Walker, P., & Bian, J. (2021). Federated learning for healthcare informatics. Journal of Healthcare Informatics Research, 5(1), 1-19. https://doi.org/10.1007/s41666-020-00082-2