2023 15th International Conference on Software, Knowledge, Information Management and Applications (SKIMA) | 979-8-3503-1655-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/SKIMA59232.2023.10387304

Unleashing the Power of Federated Learning in Fragmented Digital Healthcare Systems: A Visionary Perspective

Marzia Hoque Tania
Centre for Big Data Research in Health
University of New South Wales
Sydney, Australia, and
Institute of Biomedical Engineering
Department of Engineering Science
University of Oxford
Oxford, United Kingdom

David A. Clifton
Institute of Biomedical Engineering
Department of Engineering Science
University of Oxford
Oxford, United Kingdom, and
Oxford Centre for Advanced Research (OSCAR)
University of Oxford
Suzhou, China

***Abstract*—Digital healthcare landscape, including infrastructure, governance, interoperability, and user adoption, are continuously evolving, some taking more centralised approach, while others with higher degree of fragmentation. Attitude towards centralised healthcare systems in affluent countries are primarily influenced by historical development, infrastructure investments, and regulatory frameworks, which offers advantages with respect to standardised practises, centralised decision making, and economies of scale. In contrast, complexities due to diverse stakeholders, interoperability challenges, privacy and security concerns often pose challenges in achieving a completely centralised healthcare system even in high income countries such as the United Kingdom or in federal systems such as the United States. Moreover, decentralised healthcare systems are more prevalent in resource-poor countries. This paper presents our viewpoint and perspectives on the potential of federated learning in decentralised healthcare systems, especially in countries with infrastructure constraints and discusses its advantages, privacy and security concerns, and challenges. As data-hungry artificial intelligence-enabled systems are gradually changing the healthcare ecosystem, federated learning presents an opportunity for distributing the machine learning training process across multiple decentralised edge devices with reduced data transfer. Therefore, the decentralised digital healthcare system can leverage the collaborative model training while protecting highly sensitive and personal health information. However, challenges related to data heterogeneity, communication latency, and model aggregation need to be addressed for successful implementation of such systems. Adapting the federated learning framework to the specific needs and constraints of low and middle-income countries is crucial to unlock its potential in improving healthcare outcomes.**

***Index Terms*—Federated Learning, Decentralised Healthcare Systems, Collaborative Model Training, Data Privacy, Security, Healthcare Innovation.**

DAC was supported by an NIHR Research Professorship, an RAEng Research Chair, the InnoHK Hong Kong Centre for Cerebro-cardiovascular Health Engineering (COCHE), the NIHR Oxford Biomedical Research Centre (BRC), and the Pandemic Sciences Institute at the University of Oxford. MHT received support from the Wellcome Trust (Grant ID: B9R03140).

I. Introduction

A. Global Healthcare Systems

There are four major healthcare system models in the healthcare system typology, namely the Beveridge model (for example, United Kingdom, Cuba), the Bismarck model (for example, Germany, Japan), the National Health Insurance model (for example, Canada, South Korea) and the Out-of-Pocket model (many low and middle countries), to describe a country's healthcare system based on its different organisational, funding, and delivery approaches.

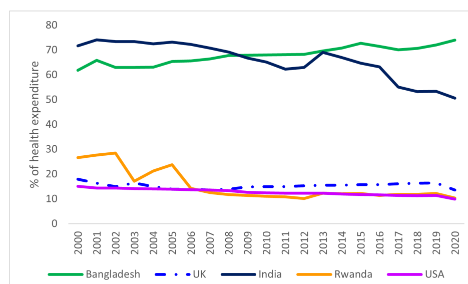

Fig. 1. Comparative Out-of-Pocket Expenditure Analysis in Diverse Healthcare Structures Using WHO Global Health Expenditure Data [1]

Centralisation in the healthcare system enables efficient coordination, quality control, and standardised practises, centralised decision-making, resource allocation, and optimisation. Implementing such an integrated national healthcare system is often influenced by historical development, infrastructure investments, and regulatory frameworks.

The adoption of centralised healthcare systems does not depend solely on the economic status of a country. Jurisdictions, existing healthcare landscape of a country, cultural norm and stakeholder diversity can result in varying or lower degree of centralisation based on unique requirements, local needs, budgets, timelines and priorities. For example, the United States, as a federal country, has a complex healthcare system, regulated by federal, state and local authorities, varying in policies on insurance coverage, healthcare delivery, patient safety, and quality of healthcare. Moreover, there is a multifaceted mix of public and private components with a wide range of healthcare providers with different ownership models and fragmented coverage. Moreover, privacy and cyber security considerations play a crucial role in centralised Electronic Health Records (EHR) implementation.

The National Health Service (NHS) in the United Kingdom is centralised in terms of funding and overarching policy, however, there are regional variations and differences in healthcare delivery practices within different parts of the UK. NHS uses a complex software structure to manage various aspects of healthcare delivery, administration, and information management. For instance, EHR systems are employed to digitally store and manage patient health records, while Picture Archiving and Communication Systems (PACS) are utilised specifically for managing medical imaging data. Hospital Information Systems (HIS) enable management of various aspects of secondary and tertiary operations, including patient registration, scheduling, admissions, billing, and inventory management in the hospital, streamlining administrative tasks and improving communication within healthcare facilities. Clinical Decision Support Systems provide healthcare professionals with evidence-based recommendations for patient care. Automation is in place for appointment scheduling and patient portals, managing prescription orders, medication administration, inventory control within NHS pharmacies, and certain information exchange or sharing across different healthcare organisations. Although the IT landscape within the NHS is dynamic and constantly evolving, complete integration of IT systems in a complex healthcare system like the NHS is a challenging process.

B. Health IT Ecosystem in Resource Constraint Countries

Limited resources and infrastructure pose challenges in implementing centralised healthcare systems in economically disadvantaged countries. Irregular fragmentation is more prevalent in such systems, relying on localised healthcare facilities, decentralised governance structures, and community health workers.

Decentralised healthcare systems in low and low-middle income countries often emphasise on community-based initiatives, and preventive care that can aid in overcoming geographical barriers, facilitates healthcare access in remote areas, and empowers local communities to address their specific healthcare needs. Leveraging existing community resources, including traditional healers and local knowledge, these systems deliver tailored healthcare services. Moreover, decentralisation proves to be adaptable to the unique challenges, enabling efficient resource allocation at a local level.

In the absence of a nationwide digital health ecosystem, decentralised systems, operating on a smaller scale and bringing healthcare closer to local communities, can still effectively reduce barriers, improve availability and patient satisfaction, optimise resources and make services more affordable and efficient. It can contribute to building more resilient healthcare facilities across regions, fostering a culture of innovation and adaptability, deploying new technologies [2], allowing for swift responses to emerging trends, and transformation of care models to meet evolving needs within the healthcare settings.

II. Privacy and Security Concerns

The evolving digital healthcare systems have gained significant attention in recent years due to its increasing potential in improving patient outcome, when coupled with Artificial Intelligence (AI) and Machine Learning (ML). AI-enabled EHR facilitates data-driven insights using accessible health records of a patient's medical journey, reduces medical errors, enhances decision making at individual as well as population level and improves quality of care. However, as Low- and Middle-Income Countries (LMICs) confront challenges in information governance due to limited resources, infrastructure, and legal frameworks, even less attention is provided to data ethics, including in healthcare, due to competing priorities, limited awareness, and cultural factors [3].

Low digital literacy in LMICs, greatly influenced by education and socioeconomic factors, is instrumental in contributing to the absence of digital trust. To heighten the lack of digital literacy, the absence of clear guidelines and ethical considerations in place can have a catastrophic consequences in healthcare settings. In addition to cyber threats, the failure to recognise data security vulnerabilities, compromised privacy and consent, and the costs associated with unauthorised access erode trust, hindering the long-term adoption of digital health services.

In high income countries (HICs), although regulations on data privacy and protection are in place, AI regulations are still an ongoing endeavor. Advancing technology and increasing global awareness on AI and data ethics are leading to more discussions and potential development of guidelines and regulations in LMICs. World Health Organisation (WHO) recently announced six principles, including protecting human autonomy, promoting well-being and safety, ensuring transparency and accountability, fostering inclusiveness and equity, and promoting AI that is responsive and sustainable [4]. However, implementing effective information governance practices remains challenging in LMICs. Therefore, fragmented, advanced digital healthcare systems can address privacy concerns by distributing data, implementing security measures, and enabling local control and access controls.

III. Federated Learning

In 2016, Google presented the concept of ML model training on distributed data sources with reduced risk of data leakage, called federated learning (FL) [5]–[7]. Later extensions concentrated on privacy-conscious collaborative learning on local devices or edge-nodes across a number of organisation using horizontal as well as vertical data partitioning based on user or device identifiers without centrally collecting and storing the data.

Imagine a situation with n data owners, each possessing datasets from d_1 to d_n. The classical ML would require to aggregate the dataset as the following before model training.

$$D_T = d_1 \cup \ldots \cup d_n$$

Whereas, FL enables all data controllers to retain their individual dataset, d_j, sharing only model-parameters accumulated from each node subsequent to local training, to update the global model at the central server.

Other decentralised methods, like Blockchain, provide secure storage in an unalterable chain of blocks linked by cryptographic hashes. Patient-centred health systems on peer-to-peer (P2P) protocols ensure direct user communication, enhancing data sovereignty but facing trust issues. However, the decentralised data often challenges traditional machine learning at the edge, promoting FL as a potential solution, especially when combined with other approaches [8].

At its core, FL adopts various methods for training and updating models across decentralised devices. Horizontal FL, for instance, necessitates a similar data structure, offering enhanced protection against data leakage as encrypted gradients are aggregated at the server. In vertical FL, concentrating on different feature spaces, encryption is in place for entity alignment and ML training, restricting collaborators to learn other's sensitive information through the process. In FL, transfer learning techniques can also be applied when there is limited sample overlap. For example, knowledge learnt from individual dataset can be transferred to the global model or, feature representations learnt from one federated dataset can be transferred into another. More advanced model enables smaller student model to learn from the soft targets or intermediate representations of a complex teacher model. Therefore, FL shows remarkable promise for healthcare applications [9], allowing local adaptation, fostering collective learning and innovation, addressing data sharing and privacy concerns, and ensuring robust performance.

FL exhibited remarkable utility during the COVID-19 pandemic. Soltan et al. introduced an FL approach to COVID-19 screening across multiple UK hospitals [10]. Using clinical data, a global model was developed and improved using federated training. The global model achieved strong predictive performance, demonstrating the potential of federated learning for healthcare applications. Another study employed FL to develop the EXAM model for predicting oxygen requirements in COVID-19 patients [9]. Data from 20 global institutes were used for training, demonstrating FL's capability to create a predictive model without data sharing. EXAM achieved strong performance, outperforming local models and enhancing generalisability. The study showcases FL's potential for collaborative and secure AI development in healthcare.

IV. Federated Learning in LMIC Context

We live in a data-rich era, where abundance of personalised data is being generated everyday. Centralised or not, increasing implementation of EHR and boom in the medical technology industry have led to a significant increase in the availability of genetic, medical and research data, which encouraged FL to penetrate healthcare sector. However, among the participatory countries in the 3rd global survey of WHO Global Observatory for eHealth, only three of the lower-middle and low-income countries reported having a national EHR system [11]. Among South Asian nations, India has developed national strategies and standards for eHealth [12], [13]. The other south Asian countries to adopt such policies are: Maldives (2011), Bhutan (2014) and Bangladesh (2015), among which the latter two are LMICs [11]. While there has been an increasing rate of adoption for eHealth, HIS and telehealth, implementation and adoption of nationwide EHRs are more restricted to higher income countries.

Bangladesh- an LMIC, has recently implemented electronic medical record (EMR) system in twelve public hospitals [14]. While EHR and EMR are often used interchangeably, EMR focuses on medical data within a single organisation, EHR offers a broader view of a patient's health history, integrating information from multiple sources.

The Management Information System (MIS), Directorate General of Health Services (DGHS), in collaboration with HISP INDIA (Society for Health Information Systems Programmes) and a private company, implemented the OpenMRS and PACS-based system in ten tertiary hospitals. While the system is not yet fully operational, its goal is to automate various aspects of health services, encompassing prescription management, medical history tracking, pathology reports, and in-patient data.

The private healthcare sector in Bangladesh is a growing part of the healthcare system (Figure 1), consisting of privately owned hospitals, clinics, and specialty centers (Table I), offering a range of medical services, and is often found in urban areas. In the absence of centralised public-funded EMR systems, the private sector, such as some private hospitals and diagnostic centres, is currently using decentralised EMR systems [15], however, more granular statistics on these EMRs are not available [16]–[18].

Examples such as International Centre for Diarrhoeal Disease Research, Bangladesh, Diabetic Association of Bangladesh, Bangladesh Institute of Research and Rehabilitation in Diabetes, Endocrine and Metabolic Disorders, United Hospital, Evercare Hospital and Square Hospital, already have automation in place. Automation for tasks such as patient management, maintaining patient records, monitoring the availability of essential drugs, and recording blood supply data are untapped sources of health data that can be utilised for the benefit of patients.

Therefore, regardless of whether a nation-wide EHR system is currently in place, such systems are drawing attractions in LMICs, and are gradually being implemented, although often in a fragmented manner [19].

FL holds promise for healthcare systems in such settings. Its privacy-preserving nature aligns with the need to protect sensitive patient data, ensuring compliance with privacy reg-

TABLE I
PRIVATE HEALTHCARE FACILITIES IN BANGLADESH

Facility Type	# Facilities
Private Hospitals	4,452
Private Clinics	1,397
Dental Clinics	839
Diagnostic Centers	10,291

ulations. FL also allows healthcare institutions in resource-limited settings to collaborate and benefit from each other's expertise without the need for extensive infrastructure upgrade or centralised data storage. By jointly training models, healthcare professionals can share insights, resulting in improved model performance and better healthcare outcomes.

AI-enabled EHR often requires annotating information such as patient diagnoses, disease progression, medication usage and treatment outcomes, establishing ground truth for supervised learning algorithms. FL introduces the potential for participants to collectively contribute in the annotation process without sharing their raw data. Moreover, using FL, healthcare organisations with limited local data can leverage pre-trained models to benefit from knowledge gained from larger and more diverse datasets.

LMICs' community-based structure can leverage FL for enhanced health monitoring at point-of-care platforms while conserving energy by reducing the need for continuous data transmission to a central server, and enabling learning and improvement even in offline scenarios.

Data serves as the fuel that powers advanced methods such as deep learning models, enabling them to learn, adapt, and improve over time. Advanced AI models in healthcare possess the capability to assimilate insights from unstructured and diverse data sources, such as clinical notes, free-text reports, and medical images. They excel in integrating this varied information, extracting meaningful knowledge from complex relationships, ultimately contributing to the improvement of health outcomes. As most LMICs struggle to reach WHO-prescribed one doctor for every thousand citizens, AI-enabled digital healthcare system can benefit from automating time-consuming and subjective tasks and FL can support LMICs in combating data scarcity challenges to use such advanced data hungry techniques.

Moreover, collaborating with HICs, LMICs can utilise advanced resources and knowledge, while HICs can improve their ML models for ethnic sub-population and advance infectious disease control efforts. Collaborative efforts of this nature possess the potential to propel advancements within the domain of medical science by addressing hitherto unanswered and under-explored clinical inquiries. Clinical ML domain can also make progress in real-time continuous learning for healthcare predictive modelling.

V. DISCUSSION

In paradigm-shifting FL approach, instead of transferring the new gold, data, the ML model is moved to the data, offering a promising method to unleash the power of dispersed digital healthcare systems. However, initiatives are required from both HICs and LMICs to address some of the challenges associated with FL.

- Data Quality and Heterogeneity: Accurate reporting is fundamentally crucial for EHR as well as ML, which proven to be more challenging in LMICs [20]. Moreover, EHRs are inherently heterogeneous in terms of formats, quality and distribution; the diversity becomes larger due to institutional variations. Adoption of common data models [21], and development and implementation of international standardisation for healthcare are needed to facilitate FL deployment. Efforts are being also made advancing FL tackling data heterogeneity, privacy, and scalability in the context of EHRs [22].
- Biases: While FL can help to combat ML's bias towards demographic groups, FL can also be a tool for bias propagation. LMICs struggle to mitigate disparities in data collection practices and ethnic disparities in HICs' EHR requires more care to handle fairness issues.
- Limited connectivity: FL in healthcare may not require frequent communication between the central server and edge nodes, it can still introduce communication overhead. Ensuring secure communication and optimising node scheduling, especially for latency-sensitive tasks is challenging for resource-contraint systems.
- Model Aggregation: FL combines model updates, parameters and information such as gradients, feature embeddings and encrypted intermediate results to train the global model. Such integration process while producing high performance can be complex. Fair representation of data in the centralised model requires both advanced integration techniques and balancing the contributions of the participating organisations.

The healthcare landscape is gradually stepping into a new era with endless possibilities with ML. Communication, EHR and AI technologies are evolving at a pace, that can allow collaborative model training across healthcare communities breaking geographical barriers while preserving patient data privacy and enhancing data access. Moreover, FL enables local adaptation of models to specific contexts, empowering healthcare providers to customise models by incorporating cultural factors, and variations in healthcare practices. With appropriate adaptations and considerations for the context of LMICs, AI-enabled federated healthcare network has the potential to revolutionise healthcare delivery, improve patient outcomes, and empower local communities in fragmented digital healthcare systems.

ACKNOWLEDGMENT

The authors would like to thank Prof. Louisa Jorm and the Centre for Big Data Research in Health for their kind support. The views expressed in this publication are those of the authors and not necessarily those of the funders.

REFERENCES

[1] Out-of-pocket expenditure as a percentage of current health expenditure, Accessed: August 14, 2023, [Online]. Available: https://databank.worldbank.org/reports.aspx?source=2&series=SH.XPD.OOPC.CH.ZS&country=.

[2] R. Knevel and K. P. Liao, "From real-world electronic health record data to real-world results using artificial intelligence," Annals of the Rheumatic Diseases, vol. 82, no. 3, pp. 306–311, 2023.

[3] J. Thomas, R. Carlson, M. Cawley, Q. Yuan, V. Fleming, and F. Yu, "The gap between technology and ethics, especially in low-and middle-income country health information systems: A bibliometric study," Stud. Health Technol. Inform, vol. 290, pp. 902–906, 2022.

[4] World Health Organization, Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva, Switzerland: World Health Organization, 2021, ISBN: 978-92-4-002921-7 (print version).

[5] J. Konečnỳ, H. B. McMahan, D. Ramage, and P. Richtárik, "Federated optimization: Distributed machine learning for on-device intelligence," arXiv preprint arXiv:1610.02527, 2016.

[6] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," arXiv preprint arXiv:1610.05492, 2016.

[7] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*, PMLR, 2017, pp. 1273–1282.

[8] J. L. Salmeron, I. Arévalo, and A. Ruiz-Celma, "Benchmarking federated strategies in peer-to-peer federated learning for biomedical data," Heliyon, 2023.

[9] I. Dayan, H. R. Roth, A. Zhong, *et al.*, "Federated learning for predicting clinical outcomes in patients with covid-19," Nature medicine, vol. 27, no. 10, pp. 1735–1743, 2021.

[10] A. A. Soltan, A. Thakur, J. Yang, *et al.*, "Scalable federated learning for emergency care using low cost microcomputing: Real-world, privacy preserving development and evaluation of a covid-19 screening test in uk hospitals," medRxiv, pp. 2023–05, 2023.

[11] W. H. Organization *et al.*, Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization, 2017.

[12] R. Sharma, A. Rohatgi, S. Jain, and D. Singh, "The ayushman bharat digital mission (abdm): Making of india's digital health story," CSI Transactions on ICT, vol. 11, no. 1, pp. 3–9, 2023.

[13] M. M. Pai, R. Ganiga, R. M. Pai, and R. K. Sinha, "Standard electronic health record (ehr) framework for indian healthcare system," Health Services and Outcomes Research Methodology, vol. 21, no. 3, pp. 339–362, 2021.

[14] A. H. M. M. Karim, Challenges and opportunity of electronic medical record system (emr) in bangladesh, Computer Science and Engineering, Dhaka, Bangladesh, Dec. 2018. [Online]. Available: http://hdl.handle.net/123456789/3299 (visited on 12/12/2018).

[15] S. Afrin and M. Arifuzzaman, "E-health in developing countries: Bangladeshi perspective," Int J Eng Adv Technol (IJEAT), vol. 9, p. 46, 2020.

[16] Bangladesh Bureau of Statistics, "Report on the survey of private healthcare institutions 2019," Statistics and Informatics Division, Ministry of Planning, Technical Report, Jan. 2021, Surveys and Studies Relating to GDP Rebasing 2015-16 Project.

[17] Bangladesh Bureau of Statistics (BBS), Bangladesh Statistics 2020. Statistics and Informatics Division (SID), Ministry of Planning, 2020.

[18] A. Hossain, R. Quaresma, and H. Rahman, "Investigating factors influencing the physicians' adoption of electronic health record (ehr) in healthcare system of bangladesh: An empirical study," International Journal of Information Management, vol. 44, pp. 76–87, 2019, ISSN: 0268-4012. DOI: https://doi.org/10.1016/j.ijinfomgt.2018.09.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0268401218304432.

[19] M. Kumar and J. Mostafa, "Electronic health records for better health in the lower-and middle-income countries: A landscape study," Library Hi Tech, vol. 38, no. 4, pp. 751–767, 2020.

[20] K. Harrison, N. Rahimi, and M. C. Danovaro-Holliday, "Factors limiting data quality in the expanded programme on immunization in low and middle-income countries: A scoping review," Vaccine, vol. 38, no. 30, pp. 4652–4663, 2020.

[21] D. W. Roblin, K. B. Rubenstein, H. M. Tavel, *et al.*, "Development of a common data model for a multisite and multiyear study of virtual visit implementation: A case study," Medical Care, vol. 61, no. 4 Suppl 1, S54, 2023.

[22] M. Subramanian, V. Rajasekar, S. VE, K. Shanmugavadivel, and P. Nandhini, "Effectiveness of decentralized federated learning algorithms in healthcare: A case study on cancer classification," Electronics, vol. 11, no. 24, p. 4117, 2022.