Federated Learning in Healthcare Informatics Research: A Literature Review

Wentao Li

The School of Biomedical Informatics |The University of Texas Health Science Center at Houston

Acknowledgment

This research is supported by Dr. Xiaoqian Jiang and reviewed by Dr. Susan Fenton at UTHealth.

Please contact the first author via email: wentao.li@uth.tmc.edu

INTRODUCTION

Due to the complexity of machine learning algorithms, the need for larger volumes of data in healthcare informatics research is unprecedentedly high (Bellazzi, 2014; Pramanik et al., 2020). Meanwhile, with the increasing anxiety about misusing healthcare-related data, regulations like Health Insurance Portability and Accountability Act (HIPAA) and General Data Protection Regulation (GDPR) have been proposed and implemented to protect individual health records. However, these regulations also bring down the usability of healthcare-related data. Thus, Federated Learning may leverage data privacy and usability by designing a privacy-preserving collaborative learning framework. A scientific review study is carefully designed and discussed to discover the current cutting-edge federated learning models in different healthcare research scenarios and the opinions of these domain experts.

METHODS

A literature search was conducted in the PubMed database on October 09, 2022, at 15:50 CDT. Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, MeSH terms include “Medical informatics”, “public health informatics”, “privacy”, and “confidentiality”; and Text words include “federated learning” and “distributed learning”. Moreover, the query of the search is shown below.

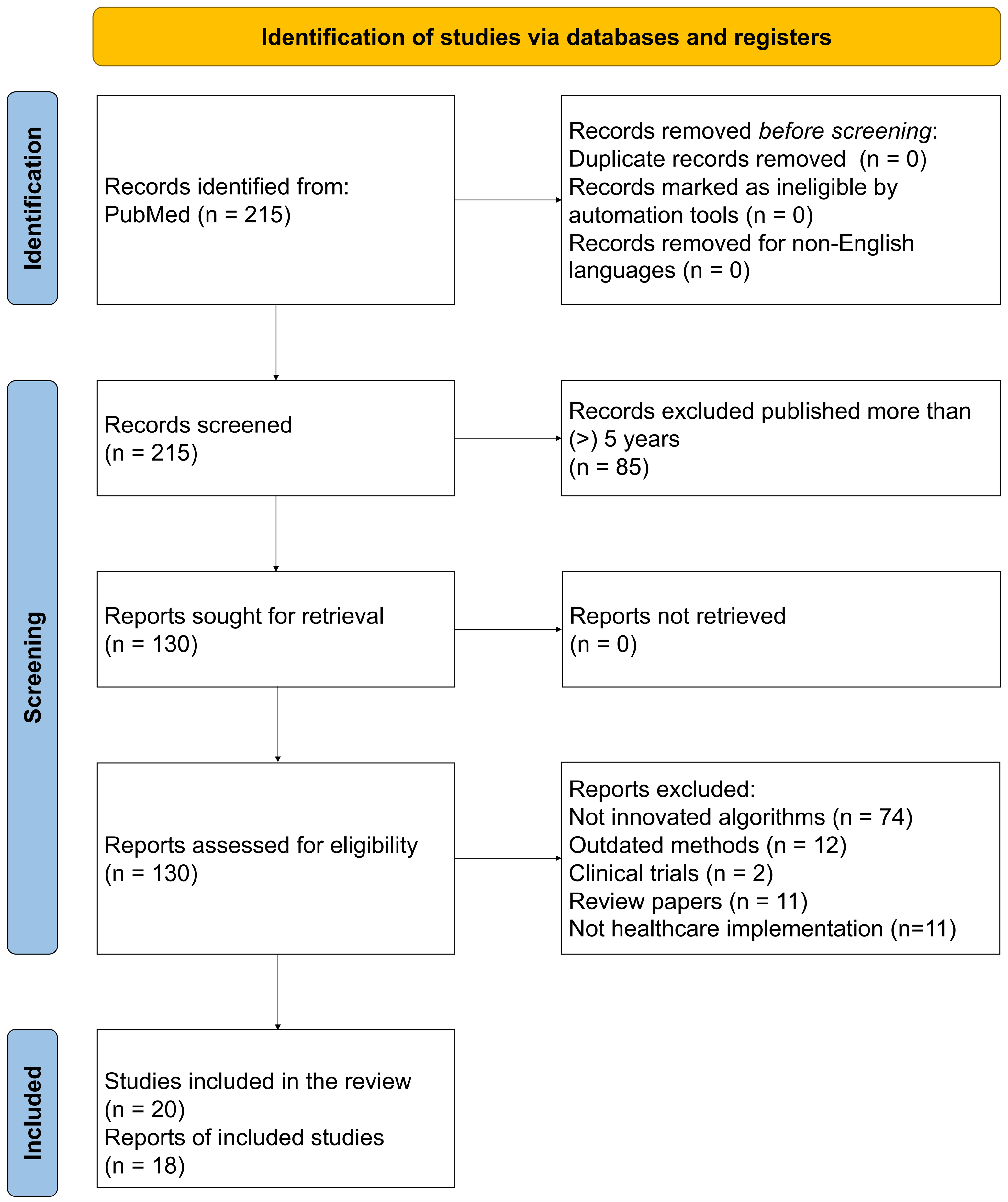
- ((((Medical Informatics[MeSH Terms]) OR (public health informatics[MeSH Terms])) AND (privacy[MeSH Terms]) OR Confidentiality[MeSH Terms]) AND ((federated learning[Text Word])) OR (distributed learning[Text Word]))

Since the study is interested to discover and summarize the cutting-edge methods published in recent years, papers were excluded if published more than five years from October 09, 2022. Thus, 85 papers were excluded. I also excluded the articles that do not have innovative algorithms and outdated methodologies, and 74 and 12 papers were excluded, respectively. Moreover, clinical trials, review papers, non-healthcare-related research, and papers in non-English languages are excluded, with a total number of 24. As a result, 20 papers were included and reviewed after the filtration. The PRISMA diagram is shown in Figure 1.

A table was created for comparison in the reviewed papers, demonstrating the details of federated learning methods.

CHART

Figure 1. PRISMA diagram for scientific review



RESULT

The following table shows the selected methods in the literature review. This study summarized the FL type, FL protection layer, implementation domain, and the strength & weaknesses of each method.

Ref	FL type	FL protection	Implementation domain	Strength	Weaknesses
(Islam et al., 2021)	Decentralized	None	Clinical data	It is a decentralized federated learning network, which is considered more secure than centralized federated learning.	It does not include a protection layer when communicating the model information
(Balachandar et al., 2020)	Centralized	None	Medical imaging data	Adopted Cyclical weight transfer (CWT) to increase performance	The model does not solve the heterogeneous problem
(Kumaresan et al., 2022)	Centralized	None	COVID-19	Developed a personalized federated multitask learning model based on the SEIR model	The method did not adopt a protection layer when training the model
(Ngo et al., 2022)	Blockchain	Hash-encrypted	Cerebellar dysfunction	Adopted blockchained FL to validate the raw data in the FL training net, which can protect the model from attacks	Blockchain technique has lower efficiency in communication
(Z. Li et al., 2019)	Centralized	Differential privacy	EHR data	Adopted differential privacy to safeguard the intermediate model information during the training	Differential privacy will sacrifice the accuracy of the prediction model
(Linardos et al., 2022)	Centralized	None	Medical imaging data	Developed a federated 3D-CNN network with a pre-trained mechanism	The method did not adopt a protection layer when training the model

CONCLUSIONS

More and more federated learning methods were adopted in healthcare informatics research. However, many still needed to apply a protection layer when transmitting model information. We are expected to see further research fixing this vulnerability.

REFERENCES

Bellazzi, R. (2014). Big data and biomedical informatics: a challenging opportunity. Yearbook of Medical Informatics, 9, 8–13.

Pramanik, M. I., Lau, R. Y. K., Azad, M. A. K., Hossain, M. S., Chowdhury, M. K. H., & Karmaker, B. K. (2020). Healthcare informatics and analytics in big data. Expert Systems with Applications, 152, 113388.

Islam, H., Mosa, A., & Famia. (2021). A Federated Mining Approach on Predicting Diabetes-Related Complications: Demonstration Using Real-World Clinical Data. AMIA ... Annual Symposium Proceedings. AMIA Symposium, 2021, 556–564.

Balachandar, N., Chang, K., Kalpathy-Cramer, J., & Rubin, D. L. (2020). Accounting for data variability in multi-institutional distributed deep learning for medical imaging. Journal of the American Medical Informatics Association : JAMIA, 27(5), 700–708. https://doi.org/10.1093/jamia/ocaa017

Kumaresan, M., Kumar, M. S., & Muthukumar, N. (2022). Analysis of mobility based COVID-19 epidemic model using Federated Multitask Learning. Mathematical Biosciences and Engineering : MBE, 19(10), 9983–10005. https://doi.org/10.3934/mbe.2022466

Li, Z., Roberts, K., Jiang, X., & Long, Q. (2019). Distributed learning from multiple EHR databases: Contextual embedding models for medical events. Journal of Biomedical Informatics, 92, 103138. https://doi.org/10.1016/j.jbi.2019.103138

Linardos, A., Kushibar, K., Walsh, S., Gkontra, P., & Lekadir, K. (2022). Federated learning for multi-center imaging diagnostics: A simulation study in cardiovascular disease. Scientific Reports, 12(1), 3551. https://doi.org/10.1038/s41598-022-07186-4

Ngo, T., Nguyen, D. C., Pathirana, P. N., Corben, L. A., Horne, M., & Szmulewicz, D. J. (2022). Blockchained Federated Learning for Privacy and Security Preservation: Practical Example of Diagnosing Cerebellar Ataxia. Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 2022, 4925–4928. https://doi.org/10.1109/EMBC48229.2022.9871371