INDO-ASIAN JOURNAL OF INFORMATION RESEARCH AND
TECHNOLOGY VOL 2 (01), JANUARY – JUNE 2025, PP. 14 - 24

ANALYSING HEALTHCARE DATA THROUGH FEDERATED LEARNING TECHNIQUES: A SYSTEMATIC LITERATURE REVIEW

Dr Mukul Rastogi
Director, Liver Transplant Fortis Hospital Noida, Uttar Pradesh, India
Email Id: Drmukul@fortishealthcare.com

ABSTRACT

The Internet of Medical Things (IOMT) and recent advancements in communication technology have revolutionized artificial intelligence (AI)-enabled smart healthcare. Due to the growing concerns about data privacy and the great scalability of contemporary healthcare networks, traditional AI techniques may not be effective in real-world healthcare scenarios where centralized data collecting and processing is required. As a new distributed collaborative AI paradigm, federated learning (FL) is especially appealing for smart healthcare since it allows numerous clients (like hospitals) to coordinate AI training without requiring them to share raw data. Consequently, we offer an extensive assessment of federated Learning's application in smart healthcare. Numerous insightful studies on Federated Learning have been published in credible publications between 2014 and 2024. The study employed a bibliometric approach, as the author discovered that it may be utilized to discern patterns in research themes, publication patterns, and citation behavior within a specific field or discipline.

Keywords*: Federated learning, Health care, cardiac data, systematic literature review, machine learning*

1. INTRODUCTION

The level of collaboration in training machine learning models across various enterprises and nations has reached an unprecedented height due to the proliferation of big data, the swift advancement of machine learning, and the growing worldwide connection (Huang et al, 2022). Data privacy concerns are the main obstacle to collaborative training in healthcare, as they restrict the sharing of data and the practical use of technically feasible interventions. As a result, privacy-preserving techniques like generative adversarial networks, blockchain technology, and federated learning (FL) are receiving more attention. It is becoming more and more popular in the medical industry as a desirable, privacy-enhancing substitute for conventional centralized training (McMahan HB et al, 2023). It has been demonstrated that federated learning is resilient to a variety of data sources, including imaging data such as magnetic resonance imaging (MRI) used to segment brain tumors, Chest X-ray for predicting the COVID-19 clinical outcome; EMR for forecasting hospitalization; colored photos, like skin photos for diagnosing skin lesions and retinal fundus photos; and histology slides for the Internet of Medical Things, genomics, and cancer diagnosis.

Federated learning has not yet been widely adopted in clinical practice, despite its benefits, and attempts are still being made to improve clinical translation. Research evaluating federated learning's robustness over a wide range of clinical domains and comparisons with current

machine learning frameworks are still under progress, in addition to federated learning being a relatively younger privacy-preserving approach (Huang et al, 2022). Furthermore, even if federated learning offers better privacy protection, sharing model updates still poses a risk to privacy. As a result, modern federated learning models incorporate additional privacy features including differential privacy (Sheller et al, 2020) and cryptographic techniques like homomorphic encryption (He et al, 2019), secure multi-party computation (Price et al, 2019), and blockchain (Wiens et al, 2019). The rest of this work is organized as follows, following a brief introduction to federated Learning and its application in the healthcare industry is described in section 2. The research approach used to carry out this study is described in Section 3. Additionally, the author presents the search results from published works in Section 4. The benchmark medical dataset, data partitioning, data distribution features, data privacy concerns, and safeguards, open challenges employed in federated learning for medical applications, and more are covered in Section 5.

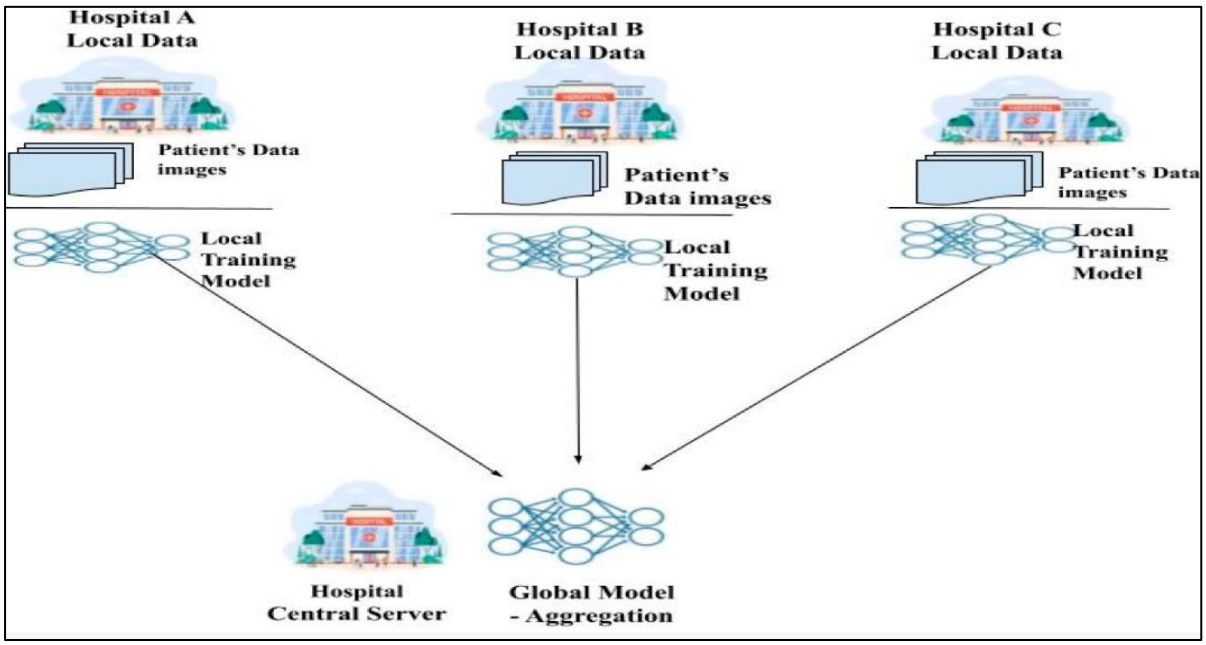
2. LITERATURE REVIEW

Federated learning, a novel approach to AI training, is quickly taking the lead in complying with several new privacy requirements. Federated learning provides a method to access the raw data coming from sensors on satellites, bridges, machines, and an increasing number of smart gadgets at home and on our bodies by processing data at their source (Zhao et al, 2018). A lot of these AI programs were trained using data that was collected and processed in one location. However, the direction of AI nowadays is moving toward decentralization. Collaboratively, on the edge, with data that never leaves your phone, laptop, or private server, new AI models are being developed. While federated learning models can be used with various machine learning methods, context, and data type are crucial. A few possible uses include the education of smartphone users, driverless cars, and wearable technology's ability to predict health hazards. Some major areas where federated learning is used are Mobile applications, Health care, Autonomous vehicles, and manufacturing-predictive maintenance.

In smart healthcare applications, deep learning technology has demonstrated encouraging outcomes in supporting clinical data-driven medical diagnosis and therapy. For example, deep learning helps with brain tumor segmentation and classification from magnetic resonance imaging (MRI) (Ge et al, 2022), text identification of medical laboratory reports (Xue et al, 2020; Harerimana et al, 2019), and cancer diagnosis and prediction (Noreen et al, 2020). A large and diversified set of training data is essential for the deep learning model to perform well on smart healthcare applications (Sultan et al, 2019). These training data came from a variety of clinical observations, including biological sensors, specific patients, hospitals, clinical settings, pharmaceutical companies, and health insurance providers. Moreover, as demonstrated in the Figure below, Zech et al. (2018) demonstrated that deep learning models trained on a single institutional dataset are susceptible to institutional data bias. Research using data from the same clinical institution has demonstrated the great accuracy of this institutional data bias. It is less effective, though, when used to data from a different organization or even within divisions within the same organization (Li et al, 2019). Concurrently, patient privacy and regulatory constraints regarding clinical data make it impractical to train deep learning models in a centralized data lake, as shown in Figure below. Therefore, to construct a single

deep learning model while protecting patient privacy and confidentiality, many healthcare institutions must collaborate to improve both the diversity and quantity of training data.

Figure 1 Federated Learning Framework



Source: Zech et al. (2018)

In the federated learning environment, there are two ways to safeguard data privacy which are perturbation and encryption. Data leakage and attackers can use these strategies. By incorporating controlled random noise into the training data or the machine learning model parameters while the model is being trained, the perturbation approach protects the privacy of both the private data and the model. In the federated learning studies published in medical applications, for example, the perturbations approaches employed include hybrid exchange parameters (Cui et al, 2021) and differential privacy (Zhang et al, 2021; Yan et al 2021). In contrast, the encryption technique, like the homomorphic encryption algorithm, protects private information and model privacy in the federated learning environment by encrypting the parameters that are shared and the gradients throughout the aggregation process (Li et al, 2019; Wu et al, 2022).

There are several challenges in the case of the adoption of federated learning some of the challenges has been described below:

- Data partitions: By combining the data of several clients, the federated Learning approach seeks to address the issue of restricted sample size when training a secure collaborative machine learning model. Selecting a horizontal or vertical data partition for federated Learning is crucial to address the issue of restricted sample size, limited sample features, or both.
- Data distribution (statistical challenge): Training data are centrally stored and balanced during training when creating a machine learning model in a centralized manner.

Federated learning, on the other hand, allowed each client to generate the training data locally, stay decentralized, and prevent access to the data of other clients.

- Data security and privacy are important concerns for medical applications. Since hundreds or maybe millions of clients are anticipated to participate, it is not feasible to assume that every client in Florida is trustworthy (Wu et al, 2022). As a result, privacy-preserving measures are required to shield medical data from unscrupulous patients or outside hackers.
- Standardized medical dataset: The quantity and quality of medical datasets have frequently hindered the creation of a strong federated Learning algorithm solution. The dataset utilized in federated Learning studies might vary widely for different types of study (Chen et al, 2022). For example, whereas some datasets concentrate on network communication performance, others concentrate on medical picture classification and segmentation performance.

Numerous insightful studies on Federated Learning have been published in credible publications between 2014 and 2024 as a result of the field's constantly evolving developments. Thus, the purpose of this study is to present a current overview of federated learning in the field of healthcare. This paper specifically outlines the Federated Learning approaches that are now in use for resolving issues with medical data as well as potential directions for future Federated Learning research in the field of healthcare applications.

3. RESEARCH METHODOLOGY

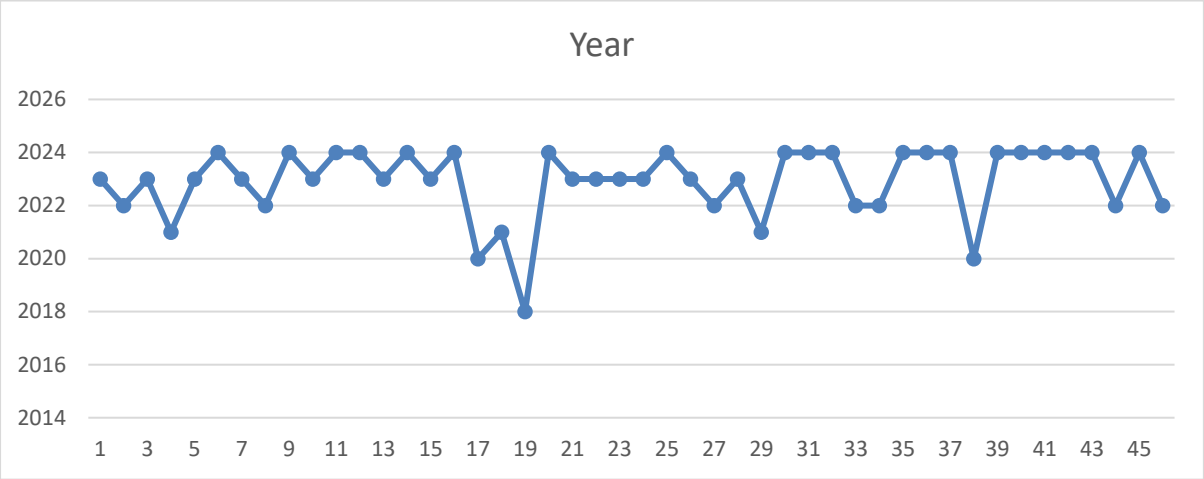
The study employed bibliometric approach, as the author discovered that it may be utilized to discern patterns in research themes, publication patterns, and citation behavior within a specific field or discipline. This can be used by researchers to find out about current knowledge gaps and new topics of interest (Paul et al, 2021). Citation counts and other bibliographic indicators can be used by researchers to assess the effect of particular papers, authors, or journals (Ellegaard and Wallin, 2015). This underlines the importance and influence of academic work. Through bibliometric study, readers can obtain a comprehensive understanding of a field. The author additionally visualized data using VOSviewer. Based on bibliometric data such as co-authorship or co-citation relationships, researchers can use VOSviewer to construct and show author, keyword, or document networks (Cheng et al., 2021). Knowing the connections and organization within a research field is made simpler by this visual representation (Zhavoronok, 2022).In this study, data has been derived from the year 2014 to 2024 from Scopus database. The keywords used for data extraction are "Federated learning" and "Health care". A total of 46 articles have been extracted the language has been filtered to English for this study.

4. DISCUSSION

Number of Articles published in a year (Scopus database)

A total number of articles are evaluated in this study where maximum articles are published after 2020. It can be seen that interest in federated learning, particularly in the healthcare area is developed among the researchers worldwide. In the last year 2023 maximum number of articles published by researchers worldwide was between 7 to 10.

Figure 2 Articles published in Scopus

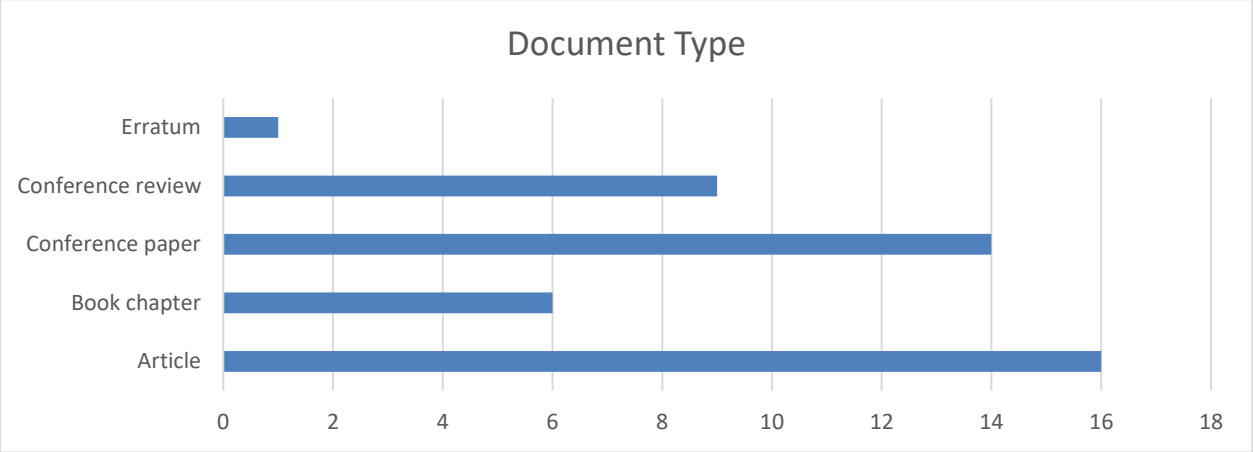


Source: Author Compilation

Types of Articles Published in Scopus

Among the researcher's conference papers is most preferred type of document where federated learning in healthcare is discussed. It can be established that federated learning in health is a prominent theme or topic in the conferences held worldwide.

Figure 3 Document type published in Scopus

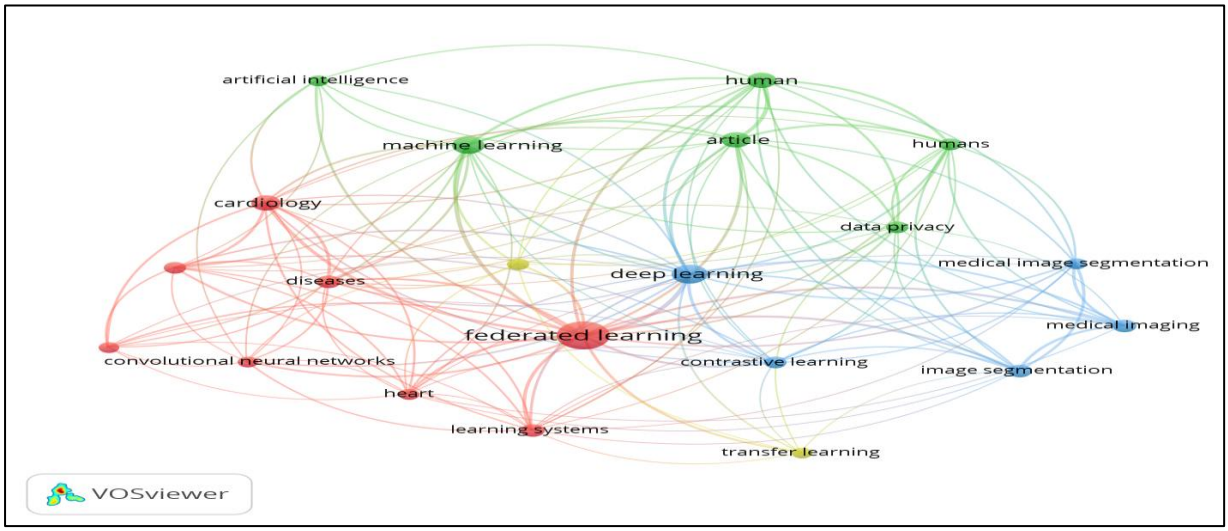


Source: Author Compilation

Keywords Co-occurrence

In bibliometric studies, co-occurrence of keywords refers to the frequency at which specific keywords appear together in academic papers within a given dataset. This approach can be used to identify themes, interdisciplinary linkages between different disciplines of study, and patterns of research subjects. In this study, the keywords most used by the various authors are artificial intelligence, cardiology, diseases, data privacy, medical, deep learning, heart, transfer learning, image segmentation, and medical imaging.

Figure 4 Keywords used by authors



Source: Author Compilation

Co-Authorship

In bibliometric studies, co-authorship analysis refers to the process of analyzing patterns of researcher collaboration through co-authored publications. Gaining an understanding of co-authorship networks can help in diffusion of knowledge, scientific productivity, and collaboration in research. One essential element of scholarly communication is co-authorship. The figure below shows the co-authorship channels. In this study also many prominent researchers working on the topic federated learning in healthcare are working closely and co-authoring many projects and articles together.

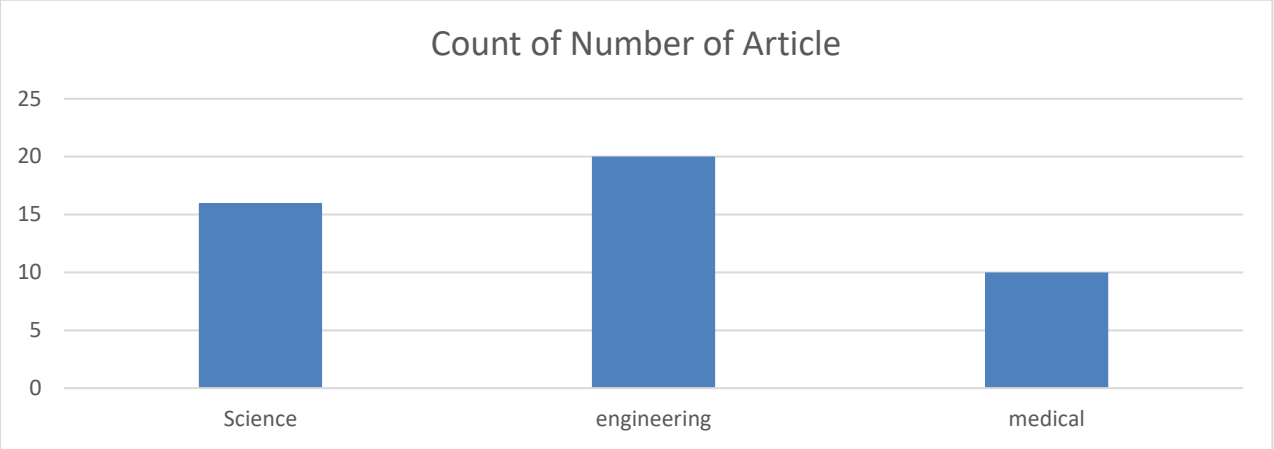
Figure 5 Co-Authorship

Source: Author Compilation

Journals most preferred by the researchers

An overview of the present status of federated Learning research in healthcare is given by this systematic review. A minority of 21.6% of the research was discovered in medical journals, while the majority were in engineering and science journals. Notably, just 5.2% of research included actual clinical application of federated Learning, indicating that federated Learning clinical use is still relatively new in the healthcare industry. This is corroborated by the exponential increase in health-related federated Learning studies throughout time Figure below, indicating a growing trend of federated Learning's appeal in the medical field.

Figure VI Journals most preferred by authors



Source: Author Compilation

Clinical application of federated Learning is not yet widely used, based on the small number of studies on its real-world clinical application. The therapeutic implementation of federated Learning faces several obstacles that need to be overcome.

Firstly, federated Learning models continue to be vulnerable to privacy violations. While federated Learning protects the privacy of raw data, sensitive health information may still be revealed through model updates shared during training. Federated Learning models are vulnerable to membership inference attacks (in which attackers use model predictions to determine if a certain individual data record was part of the training dataset) and reconstruction attacks (in which model parameters are used to infer original datasets) (Nguyenet al, 2022). Second, model performance may be harmed by the usage of non-IID health data as well as the difficulty of directly inspecting and "cleaning" in a federated Learning framework. Machine learning model performance is recognized to be significantly influenced by the quality of the data, particularly in real-world scenarios, when it comes to large, diversified, and well-labeled datasets(Nikolaidis et al, 2023). Third, to train models and share model updates, the federated Learning framework mandates that all local participating sites have access to the communication network and the requisite computational power and infrastructure(Huang et al, 2022; Liu et al, 2023). Unreliable infrastructure and networks can cause nodes to purposefully go down, and in a federated Learning scenario, the central server has no control over individual nodes.

5. CONCLUSION

In terms of popular keywords, co-authorship, number of articles published, and types of articles published on Scopus database, we showcased the progress of federated learning growth during the last ten years in the area of healthcare applications. Our ultimate goal is for this study to serve as a catalyst for further investigation on federated Learning in healthcare applications and to serve as a set of recommendations for managing private health information. We see a rise in the use of federated Learning for medical applications shortly, leading to the development of more sophisticated protocols that ensure privacy and security as well as the actual application of federated Learning technology to address pressing issues in the healthcare sector. A few things go well with this study. First of all, it offers a thorough systematic evaluation of federated Learning that includes a sizable number of papers. Second, various federated Learning models intended for the healthcare industry were discovered in the database search, which covered not only medical journals but also engineering, scientific, and general magazines. With federated Learning still in its infancy and the vast bulk of its publications being technical, this is especially crucial.

REFERENCES

Chen, Y.; Qin, X.; Wang, J.; Yu, C.; Gao, W. FedHealth: A federated transfer learning framework for wearable healthcare. *IEEE Intell. Syst.* **2020**, *35*, 83–93.

Cui, J.; Zhu, H.; Deng, H.; Chen, Z.; Liu, D. FeARH: Federated machine learning with anonymous random hybridization on electronic medical records. *J. Biomed. Inform.* **2021**, *117*, 103735.

Ge, C.; Gu, I.Y.; Jakola, A.S.; Yang, J. Enlarged training dataset by pairwise GANs for molecular-based brain tumor classification. *IEEE Access* **2020**, *8*, 22560–22570.

Harerimana, G.; Kim, J.W.; Yoo, H.; Jang, B. Deep learning for electronic health records analytics. *IEEE Access* **2019**, *7*, 101245–101259.

Huang W., Li T., Wang D., Du S., Zhang J., Huang T. Fairness and accuracy in horizontal federated learning. *Inf. Sci.* 2022;589:170–185. doi: 10.1016/j.ins.2021.12.102
He J., Baxter S.L., Xu J., Xu J., Zhou X., Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat. Med.* 2019;25:30–36. doi: 10.1038/s41591-018-0307-0.

Li, W.; Milletarì, F.; Xu, D.; Rieke, N.; Hancox, J.; Zhu, W.; Baust, M.; Cheng, Y.; Ourselin, S.; Cardoso, M.J.; et al. Privacy-preserving federated brain tumour segmentation. In *Machine Learning in Medical Imaging*; Suk, H.-I., Liu, M., Yan, P., Lian, C., Eds.; Springer: Cham, Switzerland, 2019; Volume 11861, pp. 133–141.

Liu W., He Y., Wang X., Duan Z., Liang W., Liu Y. BFG: privacy protection framework for internet of medical things based on blockchain and federated learning. *Connect. Sci.* 2023;35 doi: 10.1080/09540091.2023.2199951.

McMahan HB, Moore E, Ramage D, Hampson S, y Arcas BA. Communication-Efficient Learning of Deep Networks from Decentralized Data.arxiv Preprint at: Published online January 26, 2023. Accessed November 17, 2023.

Noreen, N.; Palaniappan, S.; Qayyum, A.; Ahmad, I.; Imran, M.; Shoaib, M. A deep learning model based on concatenation approach for the diagnosis of brain tumor. *IEEE Access* **2020**, *8*, 55135–55144.

Nguyen T.V., Dakka M.A., Diakiw S.M., VerMilyea M.D., Perugini M., Hall J.M.M., Perugini D. A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. *Sci. Rep.* 2022;12 doi: 10.1038/s41598-022-12833-x.

Nikolaidis F., Symeonides M., Trihinas D. Towards Efficient Resource Allocation for Federated Learning in Virtualized Managed Environments. *Future Internet.* 2023;15:261. doi: 10.3390/fi15080261.

Paul, J., Lim, W. M., O'Cass, A., Hao, A. W., & Bresciani, S. (2021). Scientific procedures and rationales for systematic literature reviews (SPAR-4-SLR). *International Journal of Consumer Studies*, *45*(4), O1-O16.

Price W.N., Cohen I.G. Privacy in the age of medical big data. *Nat. Med.* 2019;25:37–43. doi: 10.1038/s41591-018-0272-7.

Sultan, H.H.; Salem, N.M.; Al-Atabany, W. Multi-classification of brain tumor images using deep neural network. *IEEE Access* **2019**, *7*, 69215–69225.

Sheller M.J., Edwards B., Reina G.A., Martin J., Pati S., Kotrotsou A., Milchenko M., Xu W., Marcus D., Colen R.R., Bakas S. Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci. Rep.* 2020;10 doi: 10.1038/s41598-020-69250-1.

Wu, Q.; Chen, X.; Zhou, Z.; Zhang, J. FedHome: Cloud-edge based personalized federated learning for in-home health monitoring. *IEEE Trans. Mobile Comput.* **2020**.

Wiens J., Saria S., Sendak M., Ghassemi M., Liu V.X., Doshi-Velez F., Jung K., Heller K., Kale D., Saeed M., et al. Do no harm: a roadmap for responsible machine learning for health care. *Nat. Med.* 2019;25:1337–1340. doi: 10.1038/s41591-019-0548-6.

Xue, W.; Li, Q.; Xue, Q. Text detection and recognition for images of medical laboratory reports with a deep learning approach. *IEEE Access* **2020**, *8*, 407–416.

Yan, Z.; Wicaksana, J.; Wang, Z.; Yang, X.; Cheng, K.-T. Variation-aware federated learning with multi-source decentralized medical image data. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2615–2628

Zhang, L.; Shen, B.; Barnawi, A.; Xi, S.; Kumar, N.; Wu, Y. FedDPGAN: Federated differentially private generative adversarial networks framework for the detection of COVID-19 pneumonia. *Inf. Syst. Front.* **2021**.

Zhao, Y.; Li, M.; Lai, L.; Suda, N.; Civin, D.; Chandra, V. Federated learning with non-IID data. *arXiv* **2018**, arXiv:1806.00582.

Zech, J.R.; Badgeley, M.A.; Liu, M.; Costa, A.B.; Titano, J.J.; Oermann, E.K. Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study. *PLoS Med.* **2018**, *15*, e1002683.