ISSN: 3067-1140

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING

VOLUME: 03 ISSUE: 04

RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

Cross-Institutional Federated Healthcare Analytics

Ghatoth mishra
Independent Researcher
ghatothmishra@gmail.com

Abstract

The difficulties of working in sectors with strict privacy constraints, such as healthcare, result from the fact that data is separated and distributed across multiple sources that are unable to exchange it. The solution to this dilemma is to allow multiple data owners to collaborate on a joint task by using shared/global models while keeping the data decentralized. Three paradigms of federated machine learning are proposed: horizontal federated learning, vertical federated learning, and federated transfer learning. Federated learning processes and architectures with privacy-preserving properties that also address data provenance, access control, and interoperability.

Integrating heterogeneous distributed data sources while maintaining privacy is a major challenge for modern data and AI analysis. Sensitive data in particular, such as healthcare data, privacy is guaranteed by law. Therefore, many healthcare data sources are available based on these data sharing and data monetization are a difficult integration task. Multiple organizations are willing to collaborate by sharing insights from their private datasets, but without sharing the original datasets because healthcare data is highly sensitive. However, existing guidelines permit the use of healthcare data in Europe and America only for either research or health benefits, and not even for commercial use. Nevertheless, several companies are attempting to break this barrier by using fake-generated data for AI modeling purposes. As a result, three main criteria must be considered during the establishment of collaboration for multi-organization AI data modeling: privacy, security, and monitoring.

Keywords : Federated Learning,Privacy-Preserving Machine Learning,Secure Multi-Party Computation (SMPC),Healthcare Data Collaboration,Distributed AI Systems,Differential Privacy,Data Governance in Healthcare,Secure Model Aggregation,Cross-Institutional Data Analysis,Homomorphic Encryption,Decentralized Machine Learning,Clinical Data Sharing Frameworks,Privacy-Aware AI Architectures,Interoperable Healthcare Systems,Trustworthy AI in Healthcare.

1. Introduction

Healthcare data provide a large but underutilized resource for improving health. Security, privacy, and compliance challenges hinder sharing and federated learning can enable multi-organization collaboration without sharing data. Security concerns can also discourage participation even when there is no data leakage risk. Different federated learning architectures offer various degrees of protection against model inversion and membership inference attacks. Multi-party computation, differential privacy, and homomorphic encryption can be applied at different stages of the process to protect sensitive information, even when the model owner aims to recover it. Gaps in data governance, provenance handling, and interoperability standards represent additional risks to organisational and legal compliance. Despite these challenges, secure collaboration remains essential for developing AI models capable of responding to complex global health crises and enabling validation against real-world settings.

AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
ISSN: 3067-1140
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

Federated AI technologies can offer effective solutions for these critical security, privacy, compliance, and usability issues associated with the secure collaboration of multiple organisations in the health sector. A comprehensive overview of the state of the art was previously presented, with the focus on federated learning and secure multi-party computation paradigms for the analysis of federated healthcare data. The results of this broader analysis are reported in distinct sections addressing the well-established literature as well as recent advances.

Federated AI technologies provide promising solutions to the major security, privacy, compliance, and usability challenges that arise when multiple organizations collaborate in the healthcare sector. These technologies enable institutions to jointly analyze distributed healthcare data without requiring the direct sharing of sensitive patient information. In particular, federated learning allows machine learning models to be trained across decentralized datasets while keeping the data stored locally within each organization, thereby preserving privacy and regulatory compliance. Similarly, secure multi-party computation enables different parties to collaboratively compute results on encrypted data without revealing the underlying information to one another. A comprehensive overview of the current state of the art highlights how these approaches support secure and efficient collaboration across healthcare institutions. Previous analyses have examined both the well-established literature and the most recent advancements in federated AI technologies, providing insights into their capabilities, limitations, and practical applications. The findings from this broader review are organized into distinct sections that discuss foundational research as well as emerging developments in federated learning and secure computation techniques for federated healthcare data analysis.

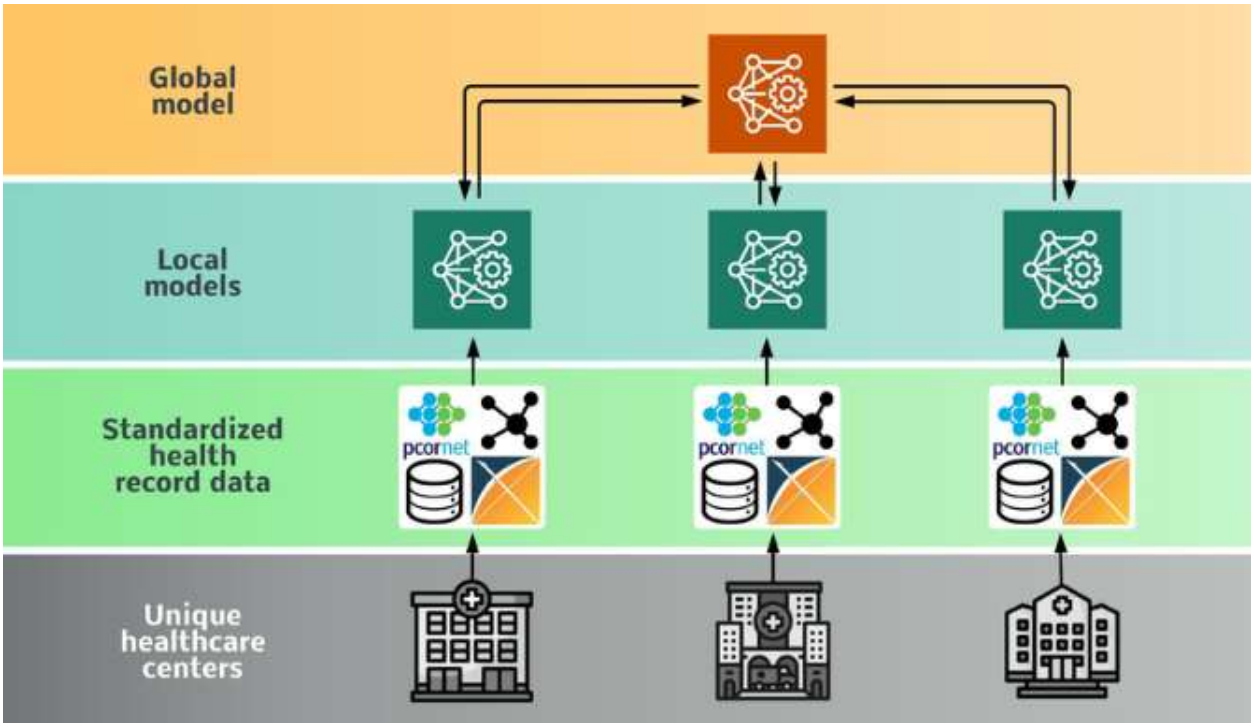


Fig 1: Federated learning model architecture

1.1. Background and Significance

In the digital world, cloud computing has become essential. It allows us to manage, analyze, and share large amounts of data without needing to create our own IT infrastructure. Yet, most data are stored in clouds owned by companies that lack appropriate security and privacy measures, raising concerns in healthcare and finance. Private cloud implementations improve security but lack the collaborative capability of public clouds. Federated learning solves these issues by keeping sensitive data at their origin and running model training directly where the data are located. Organizations collaborating in federated learning are directly involved in the learning process. Paradigms are based on the amount of shared data. In horizontal federated learning, data with the same semantics are distributed among different nodes. In vertical federated learning, the same sample applies to different feature sets at different nodes. In a federated cross-silo setting, multiple organizations join forces to exploit performance at the generalization level by increasing the amount of training data.

Although federated learning paradigms provide models capable of supporting the detection of privacy issues in data associated with health and financial applications, the standard architectures for establishing federated learning failures require thorough analysis. Most implementations use a cloud model with third-party service providers acting as centralized orchestration. These cloud nodes receive all data for training but are not governed by involved authorities and thus are exposed. Using an architecture with a centralized third-party orchestrator exposes the final model to a non-governed organization. Analysis through the lens of privacy, security, and governance is crucial. When a cloud is used as a training mechanism, known and unknown attacks flow through the architecture, and adequate solutions must be defined.

Equation 1: Federated learning (FL): objective → FedAvg update (step by step)

- There are Korganizations (hospitals / labs / insurers).
- Organization $k$has local dataset $D_k$with size $n_k = \mid D_k \mid$.
- Total samples $N = \sum_{k=1}^{K} n_k$.
- Model parameters $w \in \mathbb{R}^d$.

Let per-example loss be $\ell(w; x_i, y_i)$. Local objective:

$$F_k(w) = \frac{1}{n_k} \sum_{i \in D_k} \ell(w; x_i, y_i).$$

A standard global objective consistent with “train across decentralized datasets” is:

$$F(w) = \sum_{k=1}^{K} p_k \; F_k(w), \text{where } p_k = \frac{n_k}{N}.$$

2. Background and Motivations

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

Federated Learning (FL) is a distributed machine learning (ML) scheme allowing algorithm training on non-IID sensitive data across organizations while maintaining data privacy. Compared to traditional distributed ML or parallel ML, FL offers decentralized training without data sharing and enables edge-device-based federated transfer learning. Stand-alone FL is not sufficient for secure multi-organization data analysis; sensitive data storage and data-labeling tasks remain at risk when performed at a central organization. Orchestration-based FL emerges as a solution for these organizations.

Healthcare data analysis typically necessitates ML-based models trained on sensitive data from multiple organizations. Healthcare centers often collaborate on specific diseases, but sensitive data sharing is challenging due to unclear data ownership, lack of consent, and privacy regulations—especially in the EU, where healthcare data is considered sensitive. FL architectures for these analyses face three primary challenges: data ownership, preventive security for sensitive data storage, and orchestration governance. While requests to a central organization can be assessed in decision-making systems, FL data-assisted processes are more complex.

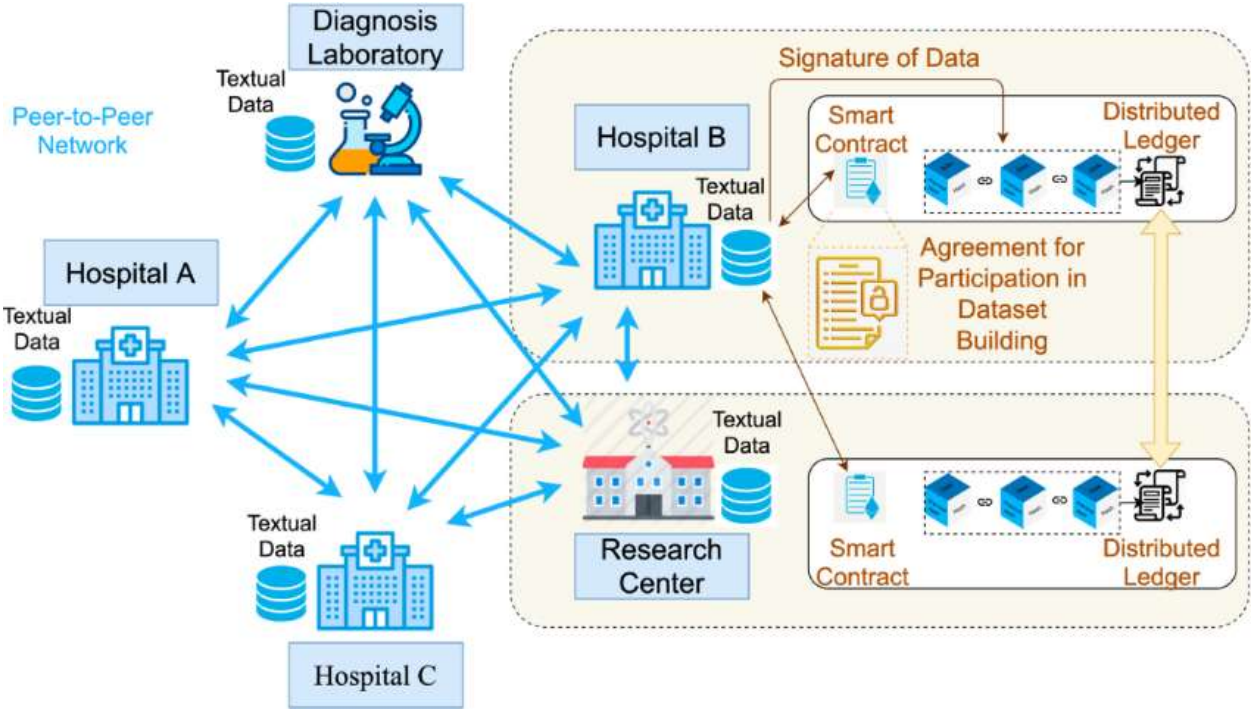


Fig 2: Background and Motivations of Federated AI

2.1. Research design

Recognizing that healthcare data contain a wealth of information needed for machine learning model training but that legal and ethical implications often make cross-organization data transfer difficult or impossible, federated learning enables organizations to collaboratively build ML models without sharing data. This large-scale systematic review identifies the current state of

federated learning paradigms and AI architectures developed for secure collaborative healthcare data analysis. The discussion focuses on security, privacy, governance, and the compliance landscape, as well as exploring advanced privacy-preserving techniques such as homomorphic encryption, secure multi-party computation, and differential protection.

The federated learning architecture reviewed allows multiple entities to collaboratively analyse their data while remaining compliant with regulations such as the European General Data Protection Regulation. Each entity holds a trusted execution environment, and deep learning models are generated by centralised model orchestration or decentralized aggregation. The federated learning paradigms used fully respect data privacy and ownership in federated environments such as the edge, cloud, and hybrid clouds, remaining robust against adversarial attacks.

Equation 2: Local SGD: from gradient descent to local training

Client $k$wants to reduce $F_k(w)$. Gradient descent step:

$$w \leftarrow w - \eta \nabla F_k(w).$$

But clients usually do **multiple local steps** E. Let $w_k^{t,0} = w^t$. For local step $e = 0, \dots, E-1$:

$$w_k^{t,e+1} = w_k^{t,e} - \eta\, g_k^{t,e},$$

where $g_k^{t,e}$is a stochastic gradient computed on a mini-batch $B_k^{t,e} \subset D_k$:

$$g_k^{t,e} = \frac{1}{|\, B_k^{t,e} \,|} \sum_{i \in B_k^{t,e}} \nabla_w\, \ell(w_k^{t,e}; x_i, y_i).$$

After Esteps, client returns $w_k^{t,E}$to server.

Define the model delta (update) from client k:

$$\Delta_k^t = w_k^{t,E} - w^t.$$

The simplest federated averaging (FedAvg) aggregates client models (or deltas) using data-size weights:

$$w^{t+1} = \sum_{k \in S_t} \frac{n_k}{\sum_{j \in S_t} n_j}\, w_k^{t,E}.$$

Equivalently in delta form:

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

$$w^{t+1} = w^t + \sum_{k \in S_t} \frac{n_k}{\sum_{j \in S_t} n_j} \Delta_k^t.$$

3. Federated Learning Paradigms for Healthcare

Federated Learning Paradigms for Healthcare. The need for multi-party cooperation in healthcare data analytics can be met using a specific variant of distributed learning called federated learning. FL enables learning a shared model across multiple organizations while preserving the data privacy within the organizations. FL is broadly categorized into horizontal FL, vertical FL, and federated transfer learning.

Horizontal federated learning (HFL) is considered when multiple parties share the same feature space (the feature sets of the local datasets are the same) but with different samples (the sample space of the local datasets is different). A typical FL setup appears in supervised learning when each organization owns a local dataset that belongs to the same feature space. The goal of HFL is to train a global model that generalizes better than any local model while distributing the learning process without sharing the local datasets. The process of HFL is typically orchestrated by an external servicer that oversees the communication of the local machines with the server and integrates the local models to create an updated global model.

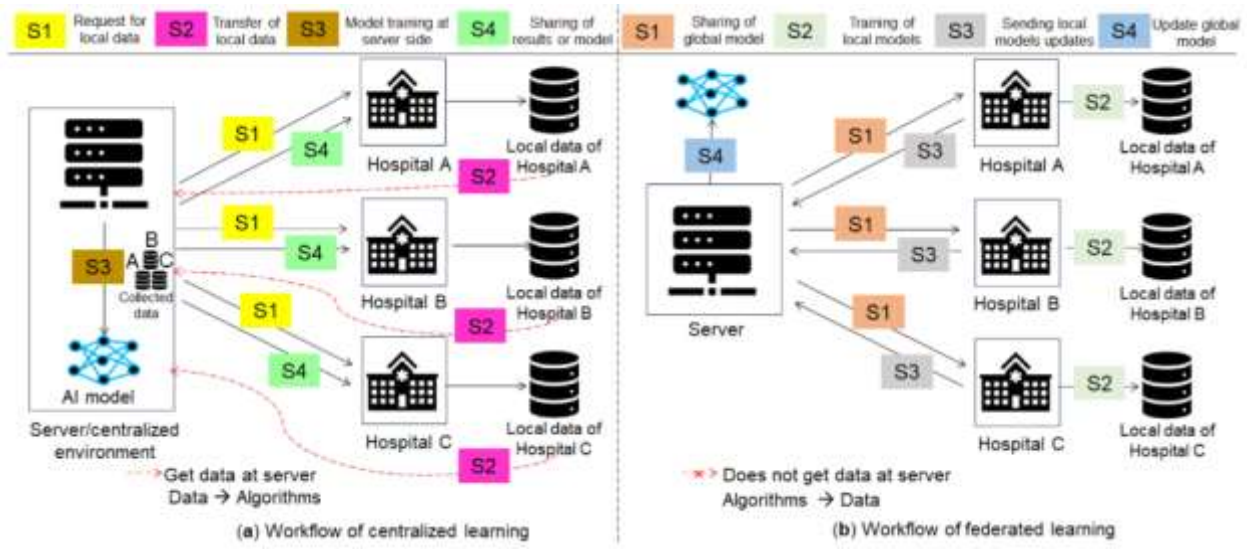


Fig 3: Analysis of Federated Learning Paradigm in Medical Domain

3.1. Horizontal Federated Learning

In horizontally partitioned data, each organization possesses samples representing the same data distribution but for different instances. A typical use case is when a federated learning system for healthcare applications is employed by a set of hospitals or clinical centers that collect samples for the same disease but represent their patients in different regions where external factors such as culture, diet, and weather may have an influence on the disease but not on the data distribution. An obvious system is the training of deep learning models for image disease classification based on data from different hospitals or clinical centers.

Two main methods can be identified for horizontally partitioned data. The first method involves sharing the weights of the model with a central server, which performs only the aggregation task. The second method is for client-sending layers or model parameters. Although the two methods generated low regional accuracy in some applications, the overall accuracy using federated learning exceeded that of independent model training. Additionally, both horizontal partitioned data methods faster required less communication time than the non-federated learning model training method. Moreover, the systems experienced greater performance with a larger number of participating hospitals.

Equation 3: Vertical FL (VFL): feature partitioning

Same patient IDs (or partially overlapping), but features split:

$$x_i = [x_i^{(A)}, x_i^{(B)}], x_i^{(A)} \in \mathbb{R}^{d_A}, x_i^{(B)} \in \mathbb{R}^{d_B}, d_A + d_B = d.$$

Model parameters also split:

$$w = [w^{(A)}, w^{(B)}].$$

Now the **prediction** often needs both parties. For a linear/logistic model, define:

$$z_i = w^\top x_i = (w^{(A)})^\top x_i^{(A)} + (w^{(B)})^\top x_i^{(B)}.$$

4. Architectural Models for Secure Collaboration

Architectural models based on a centralized orchestrator appear to be the most effective for securely sharing healthcare data between multiple organizations. In these models, data is processed by second-party organizations using cryptographic techniques and sent to a third-party organization for storage and further use. The advantage of such models is that the third party can be responsible for computing-intensive tasks, such as training federation models.

Example architectures belonging to this category include SecureFL, which incorporates remodels of semi-honest third-party organizations and training malfeasant third-party organizations, Fed-implantable, which aims to cater to the needs of FDHC systems supported by a centralized third-party organization, and ProFL, which not only protects data privacy but also safeguards model confidentiality against the data provider. Analyzing a group of healthcare organizations under common control and

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

regulation, a cross-domain federated softmax learning framework is proposed that considers both interaction and privacy-preserving requirements.

Another variant uses federated adaptive random selection to dynamically select training sites and hide their identity by secret sharing and local data perturbation. The model overcomes the inefficiency of the passive semi-honest third party and Byzantine-resilient heavy gradients in subsequent faceted learning. The proposed faceted deep learning framework ensures that patients can actively control, select, and share personal sensitive data with specific hospitals.

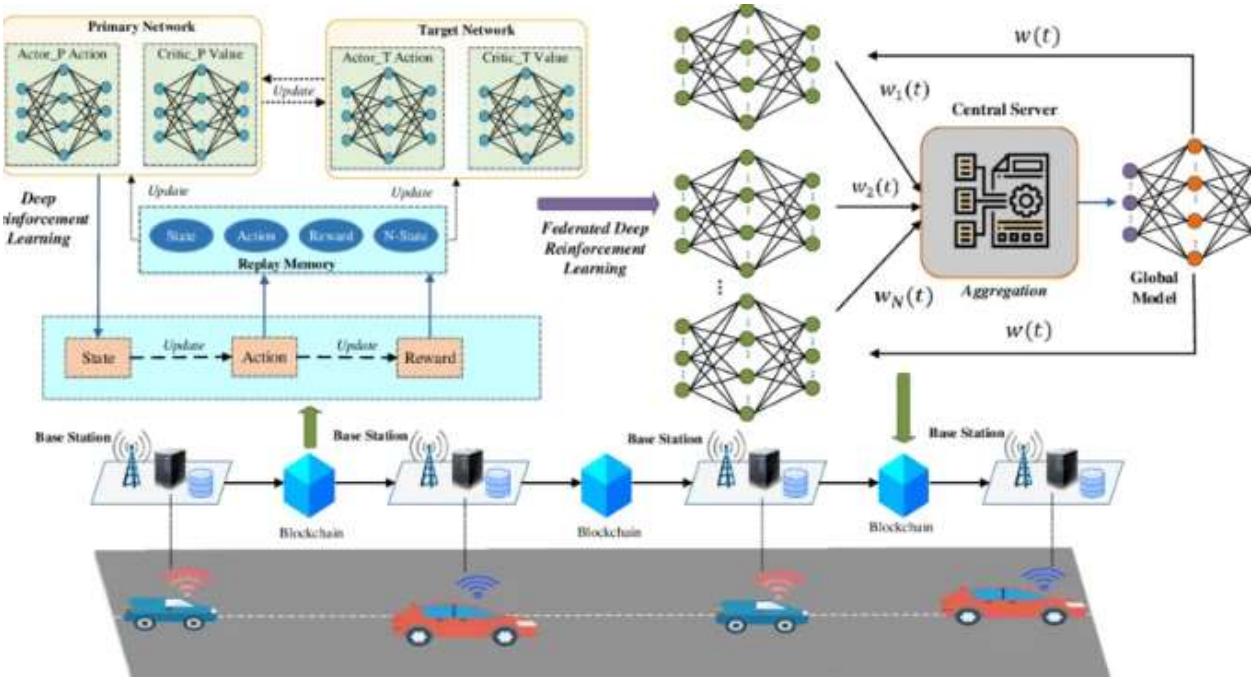


Fig 4: The architecture of federated learning

4.1. Centralized Orchestrator Architectures

An architecturally centralized federated learning model helps multi-organizational entities concerned with patient data analyses or private ML/AI solution offerings. The parties such as health providers, medical device manufacturers, health and pharmaceutical, insurance companies, supervised by a regulatory organization for the conducts have their own distributed dataset confined to their premises due to jurisdiction regulations and data-usage contracts. A central orchestrator allows independent multi-organization entities to participate in a federated paradigm that is centrally scrutinized and monitored, addressing some of the safety and reliability requirements needed for federated learning, such as patient safety, legal compliance of patient data usage, data governance, execution control, security, risk control, and thedata being the patients' assets.

An institution not gaining any benefits from the training, testing, or validation of services or products allowed to be use or offered with patient data can supervise the data-access contracts & processes of the data-consuming parties, the results gained from the analyses, and their distribution. The use of patient data by unauthorized parties or beyond the restrictions will be detected and reported by the supervisor. Once the usage of patient-data by these organizations is established, other institutions can readily join the federation, sharing & using ML solutions already trained by other agents in the association. The presence of one of the multi-organization sides (particularly health providers) as agent practicing model distinction is needed for patient safety during the supervised multi-organization federated analysis.

Equation 4: Vertical FL example: logistic regression gradients with secure sum (full derivation)

Assume binary logistic regression, label $y_i \in \{0,1\}$, probability:

$$\hat{y}_i = \sigma(z_i) = \frac{1}{1 + e^{-z_i}}.$$

Cross-entropy loss:

$$\ell_i(w) = -y_i \log(\hat{y}_i) - (1 - y_i)\log(1 - \hat{y}_i).$$

A standard result (derive via chain rule) is:

$$\frac{\partial \ell_i}{\partial z_i} = \hat{y}_i - y_i.$$

Because $z_i = (w^{(A)})^\top x_i^{(A)} + (w^{(B)})^\top x_i^{(B)}$,

$$\nabla_{w^{(A)}} \ell_i = \frac{\partial \ell_i}{\partial z_i} \cdot \frac{\partial z_i}{\partial w^{(A)}} = (\hat{y}_i - y_i)\, x_i^{(A)}.$$

Similarly,

$$\nabla_{w^{(B)}} \ell_i = (\hat{y}_i - y_i)\, x_i^{(B)}.$$

5. Privacy, Security, and Compliance Considerations

Specific privacy or security techniques used to protect model training depend on the selected federated learning paradigm and the sensitivity of the data involved. Healthcare data is by default sensitive, requiring careful consideration of both data privacy and model integrity to ensure compliance with applicable regulations. At a conceptual level, the relevant approach is based on data minimization: sharing models rather than the underlying data violates neither security nor privacy.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

Privacy provisioning techniques can be classified into three categories: model level, communication level, and data level [59]. Techniques acting at the model level can be roughly categorized according to whether privacy or security of the model itself is considered. Privacy-preserving data sharing, privacy-preserving data mining, and privacy-preserving machine learning belong to the former class. Homomorphic encryption (HE) and secure multi-party computation (SMPC) are typical techniques for the latter task. Techniques acting at the communication level aim to protect model updates; homomorphic encryption, differential privacy (DP), secret sharing, and model perturbation are all widely applied techniques. Finally, data-level techniques focus on the data used for training, and HE, SMPC, DP, feature selection, mulit-party computation, data encryption, and data perturbation are typical candidates.

Homomorphic encryption, secure multi-party computation, and differential privacy are now treated in more detail to illustrate the specifics of the techniques involved. Further information can be found in survey articles on privacy-preserving techniques for federated learning.

Equation 5: Secure sum via additive homomorphism (HE) (equation)

Federated AI Architectures for …

If Enc is additively homomorphic:

$$\mathrm{Enc}(a) \oplus \mathrm{Enc}(b) = \mathrm{Enc}(a+b).$$

Protocol sketch:

1. Party A computes $z_i^{(A)}$, sends $\mathrm{Enc}(z_i^{(A)})$to B.
2. Party B computes $z_i^{(B)}$and combines:

$$\mathrm{Enc}(z_i) = \mathrm{Enc}(z_i^{(A)}) \oplus \mathrm{Enc}(z_i^{(B)}) = \mathrm{Enc}(z_i^{(A)} + z_i^{(B)}).$$

3. Authorized decryptor gets z_i, computes $\hat{y}_i$, returns only what's necessary (often masked/limited) for gradient steps.

The paper also lists SMPC.

Federated AI Architectures for …

Simple additive secret sharing over a large modulus q:

- A secret $s$is split into shares $s_1, \ldots, s_m$such that

$$s \equiv \sum_{j=1}^{m} s_j \ (\mathrm{mod} q).$$

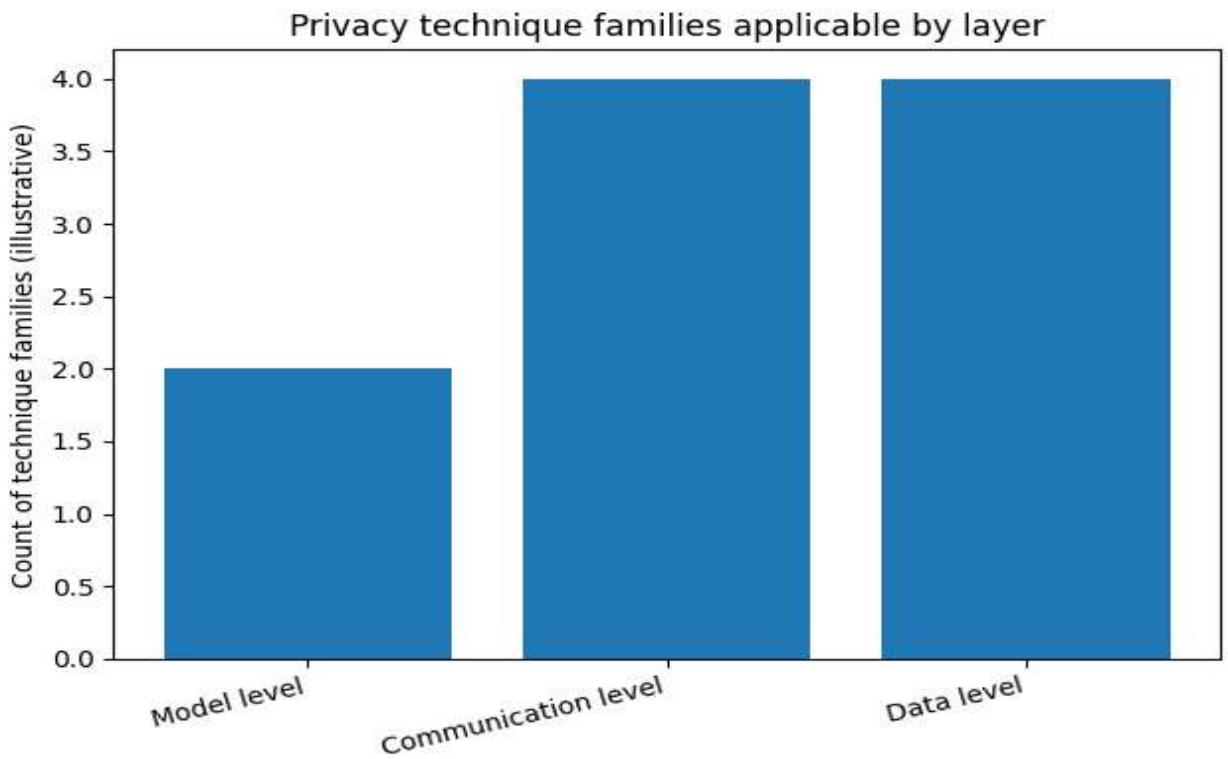
ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

5.1. Data Privacy Techniques (HE, SMPC, DP)

HE, SMPC, and differential privacy (DP) are the main data privacy techniques integrated into machine learning algorithms in real-world applications for secure data analysis in federated learning systems. HE empowers partner organizations to keep sensitive data private and encrypted. It allows computing on ciphertexts, generating an encrypted result that remains unintelligible until decryption. Thus, organizations can retain the confidentiality of valuable sensitive data in the unique or extended database during the federated learning process.

SMPC enables two parties to jointly evaluate a function while keeping their inputs private. Distributing the data across partner organizations and processing data without sharing it is a powerful approach to solving data-sharing and data-sensitivity issues. Differential privacy is achieved by adding noise to the data set in an appropriate way and removing it when the final result is obtained. When the outputs are published, neither the mechanism nor the statistician can infer anything about the input.

Data privacy techniques are critical in horizontal and vertical federated learning. However, a detailed discussion of the benefits, advantages, and limitations of such techniques in vertical or multi-party FL is beyond the scope of this work.



ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

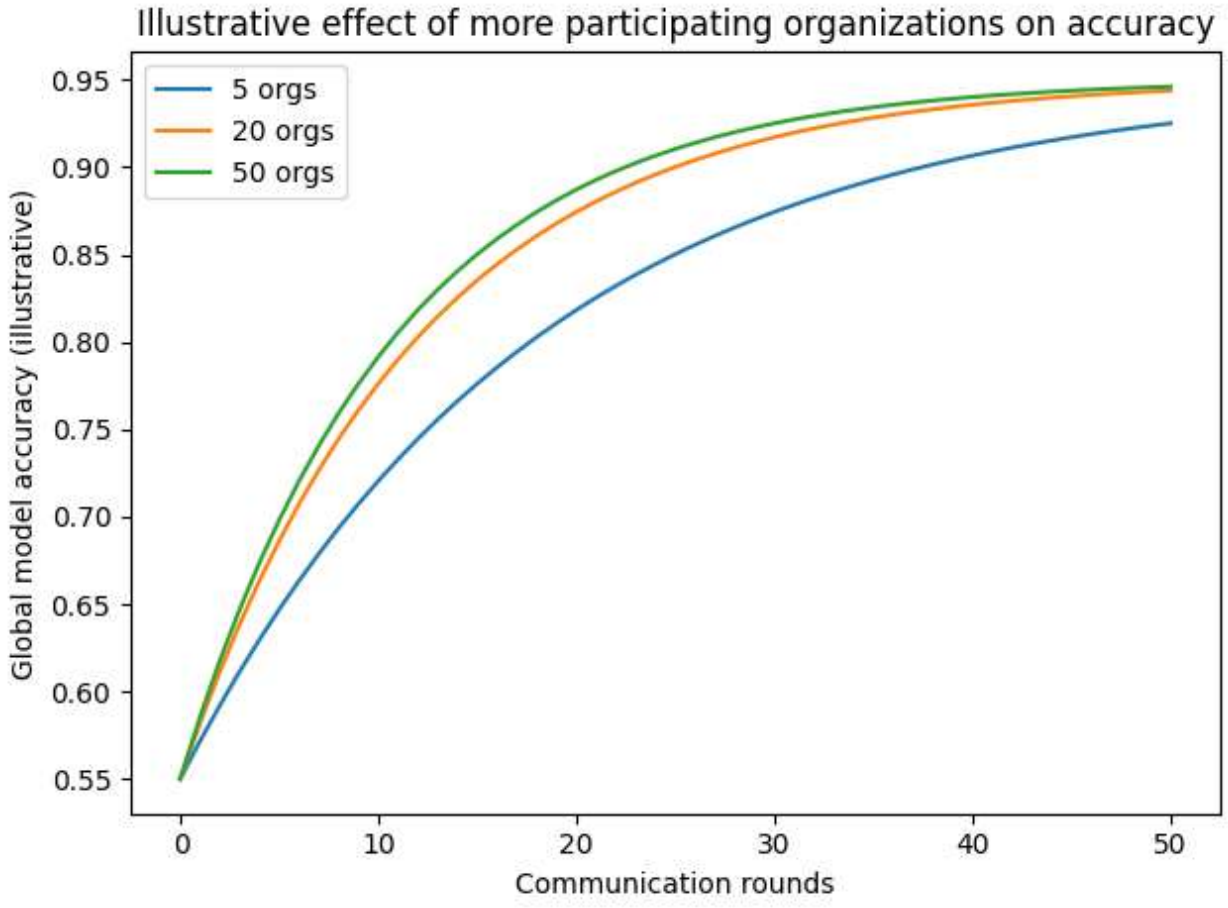
6. Data Governance, Interoperability, and Standards

Data governance aims to answer crucial questions regarding data management, such as possession and acceptable usage, including techniques for auditability, provenance, and access control. These aspects are especially relevant when the data owner is not the user, a typical situation in data-sharing or federated settings. Consequently, data governance is directly related to data privacy and its regulations: without provenance, auditability, and defined usage policies regarding access and trade, it is difficult, if not impossible, to comply with data privacy rules.

Data interoperability allows different systems, organizations, and people to exchange and mutually use data. It especially highlights the syntactic and semantic interoperability enabled by the use of common frameworks, protocols, and standards in the definition of data description and access templates. Nonetheless, achieving semantic interoperability still requires additional effort, even with common definitions, due to differences in the data models used in data sources. Such differences may arise from the underlying technologies (for example, relational databases and blockchain), the specific domain (for example, healthcare and finance), or heterogeneous analysis perspectives (for example, prediction and what-if scenarios). While these concerns often do not apply in the same organization, they emerge when data are shared with third parties. Finally, data standards constitute a necessary condition—not sufficient, though—for data interoperability. Without common frameworks and protocols, achieving interoperability across different domains requires customizing connections for each pair of data sources.

Data interoperability refers to the ability of different systems, organizations, and individuals to exchange data and use it effectively across various platforms. It relies heavily on **syntactic and semantic interoperability**, which are supported by shared frameworks, protocols, and data standards that define how data is structured and accessed. However, achieving full semantic interoperability remains challenging because data sources often rely on different **data models and technologies**, such as relational databases or blockchain systems. Variations may also arise from differences in application domains—like healthcare or finance—or from diverse analytical approaches, such as predictive modeling or scenario analysis. While these issues may be less prominent within a single organization, they become more significant when data is exchanged with external parties. Therefore, although data standards are essential for enabling interoperability, they alone are not sufficient. Without common frameworks and protocols, organizations must create customized integrations between each pair of data sources, making cross-domain data sharing more complex and resource-intensive.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10



6.1. Data Provenance and Access Control

Data provenance generated by the collaboration in federated learning demonstrates the quality and trustworthiness of the trained models beyond model performances. Provenance can be used to identify the participating parties and the data used to generate the models. By exposing sufficiently detailed provenance information, the participating organizations can share their knowledge–i.e., the knowledge about what is being modeled and the data provenance–with any external users of the models, for example, regulatory organizations. Policy-based provenance access control can restrict any sensitive provenance information from being exposed to untrusted users. In healthcare scenarios, provenance information can be sensitive because it provides knowledge about the data hosting organizations and their patient populations, which potential attackers can use for future data breaches, for example, identity theft.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

In the cyber security domain, the copyright of the used data and who provided the data for modeling can be extremely sensitive. Therefore, defining a security model for provenance access control can help reduce the risk of exposing sensitive provenance information to unauthorized users.

7. Conclusion

Great efforts have been made to explore FL in the sensitive domain of healthcare, which has huge potential for a wide variety of real use cases. However, almost all approaches consider a single organization setting, while in practice, many use cases require collaboration among multiple organizations to gain insights. Therefore, recent studies have provided federated FL paradigms for integrating FL with other distributed learning paradigms. Still, security, privacy, and governance aspects of multi-organizations collaboration remain largely unexplored. The proposed discussion has focused on these aspects while building a design strategy for the architectures of multi-organizations collaboration. The discussion is pivotal to initiate an in-depth analysis of multi-organization FL frameworks in the sensitive healthcare domain. Several critical aspects to achieve secure multi-organization collaboration using FL have been investigated, including multiple federated learning paradigms, architectural models of secure collaboration, privacy techniques on sensitive data, and data governance for enabling and sustaining collaboration.

The progress of FL is fast, but it is still in its infancy for a multi-organization setting. Many new FL paradigms have been introduced for enabling collaboration among multiple organizations. The concept of a central operator has also been extended to permit the sharing of local models while removing dependence on a central operator. Even with these substantial developments, security and privacy aspects of collaboration at the organizational level have rarely been studied, including clarifying the notion of control over the shared sensitive data and models. In particular, there has been little interaction with privacy techniques such as HE and SMPC, which are mainstream in secure data sharing at the organization level. No exploration has taken place into governance issues related to data provenance, control over sharing during actual FHE, and the relationships with traditional data governance topics in the multi-organization context.

Paradigm	**What is shared?**	**Data partitioning**	**Typical healthcare example**
Horizontal FL (HFL)	Model updates / weights	Same features, different samples	Multi-hospital imaging classifier
Vertical FL (VFL)	Encrypted gradients / partial embeddings	Same samples (overlap), different features	Hospital + insurer joint risk model
Federated Transfer Learning (FTL)	Shared representations / transferred parameters	Different samples and features (limited overlap)	Rare disease model across countries

Table : Federated learning paradigms (summary table)

7.1. Future Trends

Key enabling technologies for future healthcare service delivery, along with government policies and funding, are likely to accelerate market development. AI-powered solutions can assist users in identifying, anticipating, and resolving potential issues regarding health and telemedicine services. Better data sharing and collaboration among European healthcare institutions will also enhance the personalization of various preventive measures. For example, using federated learning makes it possible to train cluster models that better inform personalized care interactions with patients and improve the recommendation of personalized prevention actions.

Changes in the governance of current care delivery systems will also stimulate growth. In the long term, the COVID-19 pandemic is likely to improve risk management processes around healthcare systems worldwide. All nations are expected to invest in systems designed to ensure rapid, real-time data collection and analysis in order to facilitate timely decision-making and increase liquid capital. More accessible data structures will enable integrated operations across different systems and countries. New advanced analytics and predictive models, including AI-driven solutions, can also enhance user monitoring, allowing both patients and insurance companies to act preemptively and effectively prevent risk situations.

8. References

1. Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., Liu, A., Costa, A. B., Wood, B. J., Tsai, C.-S., & others. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27, 1735–1743.
2. Sadilek, A., Liu, L., Nguyen, D., Kamruzzaman, M., Serghiou, S., Rader, B., Ingerman, A., Mellem, S., Kairouz, P., Nsoesie, E. O., MacFarlane, J., Vullikanti, A., Marathe, M., Eastham, P., Brownstein, J. S., Aguera y Arcas, B., Howell, M. D., & Hernandez, J. (2021). Privacy-first health research with federated learning. npj Digital Medicine, 4, 132.
3. Warnat-Herresthal, S., Schultze, H., Shastry, K. L., Manamohan, S., Mukherjee, S., Garg, V., Sarveswara, R., et al. (2021). Swarm Learning for decentralized and confidential clinical machine learning. Nature, 594, 265–270.
4. Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2021). The future of digital health with federated learning. npj Digital Medicine, 3, 119.
5. Kolla, S. H., & Peddi, R. K. (2024). Designing Governance-Aligned GenAI Pipelines Using Small Language Models for Enterprise Workflow Intelligence. International Journal of Science, Research and Technology, 7(6), 13256-13268.

6. Pati, S., Baid, U., Edwards, B., Sheller, M. J., Wang, S.-H., Reina, G. A., Foley, W. D., Gruzdev, A., Tannenbaum, A., Thakur, S. P., et al. (2021). Federated learning enables big data for rare cancer boundary detection. Nature Communications, 12, 3020.
7. FAMHE research team. (2021). Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. Nature Communications, 12, 5910.
8. Guo, P., Wang, P., Zhou, J., Jiang, S., & Patel, V. M. (2021). Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2423–2432.
9. Glicksberg, B. S., Celi, L. A., & others. (2021). Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: Machine learning approach. JMIR Medical Informatics, 9(1), e24207.
10. Crowson, M. G., Moukheiber, D., Robles Arévalo, A., Lam, B. D., Mantena, S., Rana, A., Goss, D., Bates, D. W., & Celi, L. A. (2022). A systematic review of federated learning applications for biomedical data. PLOS Digital Health, 1(5), e0000033.
11. Adnan, M., Kalra, S., Cresswell, J. C., Taylor, G. W., & Tizhoosh, H. R. (2022). Federated learning and differential privacy for medical image analysis. Scientific Reports, 12, 1953.
12. 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, 3551.
13. Mattaparthi, R. (2024). Transformer-Based Fault Diagnosis for Large-Scale Standby Power Generators: Partial Discharge Pattern Recognition at Hyperscale Data Center Installations. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 8781-8799.
14. Nguyen, T. V., Dakka, M. A., Diakiw, S. M., VerMilyea, M. D., Perugini, M., Hall, J. M. M., & Perugini, D. (2022). A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical data. Scientific Reports, 12, 8888.
15. Wang, W., Li, X., Qiu, X., Zhang, X., Brusic, V., & Zhao, J. (2023). A privacy preserving framework for federated learning in smart healthcare systems. Information Processing & Management, 60(1), 103167.

ISSN: 3067-1140
AMERICAN ONLINE JOURNAL OF SCIENCE AND ENGINEERING
VOLUME: 03 ISSUE: 04
RECEIVED: OCTOBER 27 REVISED: NOVEMBER 06 ACCEPTED: NOVEMBER 24 PUBLISHED: DECEMBER 10

16. Rajendran, S., et al. (2023). Data heterogeneity in federated learning with electronic health records: Case studies of risk prediction for acute kidney injury and sepsis diseases in critical care. PLOS Digital Health, 2(3), e0000117.
17. Zhao, L., & Huang, J. (2023). A distribution information sharing federated learning approach for medical image data. Complex & Intelligent Systems, 9, 5625–5636.
18. Rajagopal, A., Redekop, E., et al. (2023). Federated learning with research prototypes: Application to multi-center MRI-based detection of prostate cancer with diverse histopathology. Academic Radiology, 30(4), 644–657.
19. Ma, B., Feng, Y., Chen, G., Li, C., & Xia, Y. (2023). Federated adaptive reweighting for medical image classification. Pattern Recognition, 144, 109880.
20. Reddy, V. A. R. (2024). Generative Intelligence for Healthcare Claims Processing and Personalized Benefits Management. International Journal of Advanced Engineering Science and Information Technology (IJAESIT), 7(3), 14099.
21. Liu, Z., Wu, F., Wang, Y., Yang, M., & Pan, X. (2023). FedCL: Federated contrastive learning for multi-center medical image classification. Pattern Recognition, 143, 109739.
22. Sandhu, S. S., Gorji, H. T., Tavakolian, P., Tavakolian, K., & Akhbardeh, A. (2023). Medical imaging applications of federated learning. Diagnostics, 13(19), 3140.
23. Rehman, M. H. U., Pinaya, W. H. L., Nachev, P., & Teo, J. T. (2023). Federated learning for medical imaging radiology. The British Journal of Radiology, 96(1150), 20220890.
24. Noman, A. A., Rahaman, M., Pranto, T. H., & Rahman, R. M. (2023). Blockchain for medical collaboration: A federated learning-based approach for multi-class respiratory disease classification. Healthcare Analytics, 3, 100135.
25. Nagaraj, D., Khandelwal, P., Steyaert, S., & Gevaert, O. (2023). Augmenting digital twins with federated learning in medicine. The Lancet Digital Health, 5(5), e251–e253.
26. Aravind, R. (2024). Integrating controller area network (CAN) with cloud-based data storage solutions for improved vehicle diagnostics using AI. Educational Administration: Theory and Practice, 30(1), 992-1005.
27. Guan, H., Yap, P.-T., Bozoki, A., & Liu, M. (2024). Federated learning for medical image analysis: A survey. Pattern Recognition, 151, 110424.
28. Pati, S., Kumar, S., Varma, A., et al. (2024). Privacy preservation for federated learning in health care. Patterns, 5(7), 100974.

29. Soltan, A. A. S., Thakur, A., Yang, J., Chauhan, A., D'Cruz, L. G., Dickson, P., Soltan, M. A., Thickett, D. R., Eyre, D. W., Zhu, T., & Clifton, D. A. (2024). A scalable federated learning solution for secondary care using low-cost microcomputing: Privacy-preserving development and evaluation of a COVID-19 screening test in UK hospitals. The Lancet Digital Health, 6(2), e93–e104.