International Journal of Artificial Intelligence, Data Science, and Machine Learning
Grace Horizon Publication | Volume 6, Issue 2, 64 -71, 2025
ISSN: 3050-9262 | https://doi.org/10.63282/3050-9262.IJAIDSML-V6I2P107

Original Article

Federated Learning in Medical AI: Advancing Privacy-Preserving Data Sharing for Collaborative Healthcare Research

Sriharsha Daram
Senior AWS Full stack Engineer, CGI, USA.

Received On: 02/03/2025 ***Revised On: 12/03/2025*** ***Accepted On: 28/03/2025*** ***Published On: 17/04/2025***

Abstract: *Federated Learning (FL) has emerged as a practical approach to training machine learning models collaboratively across multiple institutions, especially in domains like healthcare where patient data is highly sensitive. By allowing data to remain local while only model updates are shared, FL addresses a critical balance between innovation and privacy. This paper explores FL's growing relevance in medical AI particularly its role in improving diagnostic models, patient management, and regulatory compliance. Key contributions include a breakdown of FL's interaction with healthcare systems, a look at privacy-preserving techniques like differential privacy and homomorphic encryption, and real-world use cases in oncology, cardiology, and radiology. We present experimental results, challenges with interoperability, and a vision for FL's evolution in secure global healthcare collaboration.*

Keywords: *Federated Learning, Medical AI, Differential Privacy, Secure Aggregation, Homomorphic Encryption.*

1. Introduction

1.1 Emergence of Federated Learning in Medical AI

The integration of Federated Learning (FL) into medical AI reflects a timely response to the sector's dual need for innovation and privacy. As medical institutions grapple with enormous volumes of patient data, FL offers a unique way to build powerful predictive models without compromising data security. Rather than transferring raw data, which can raise ethical and legal concerns, institutions can train models locally and contribute to a shared global model. This shift has sparked meaningful progress in medical AI, allowing distributed learning to flourish even under strict data protection laws.

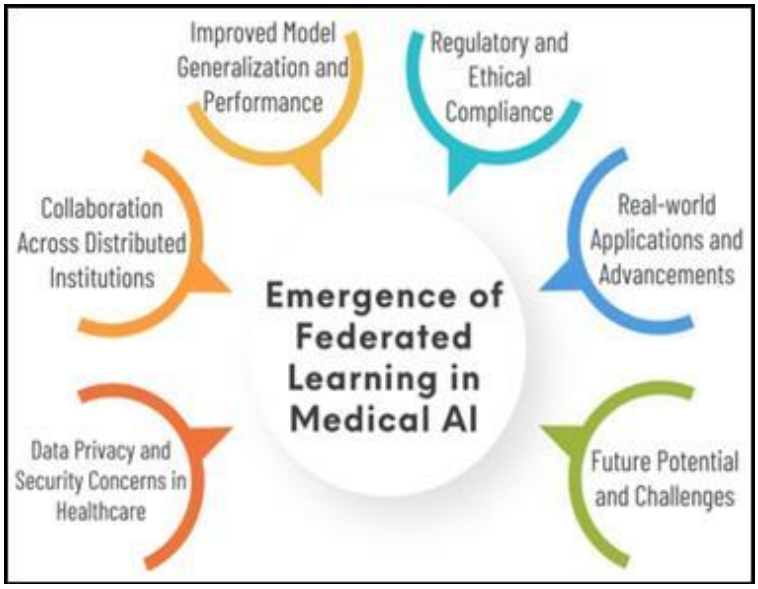

Figure 1: Emergence of Federated Learning in Medical AI

- **Data Privacy and Security Concerns in Healthcare:** The main factor that has made Federated Learning widely implemented in medical AI is the issue of data security. These ideas are rather cosy because data in health care is private, and owing to the large data sets that are usual in machine learning models, one of the biggest concerns is data leaks and user information fraud.

Federated Learning deals with this challenge so that the data is never transmitted to the central server. Still, only updates are sent to add to the aggregated update from the central server. This way, patients' privacy remains protected, but the data can be shared between institutions to enhance the creation of a better generalized AI.

- **Collaboration Across Distributed Institutions:** AI in the medical space usually requires a large and varied amount of patient data to create models that can be used on various patients with various diseases. Unfortunately, the institutions that operate in the realm of medicine are bound by legal and ethical guidelines that do not allow for the exchange of information. FL assists integration in different facilities, such as hospitals, clinics, and research institutions. FL enables the training of models using the local data from the institutions and includes only reporting model parameters like weights and gradients that can be exchanged safely between institutions. This is especially crucial in areas like medical imaging, genomics, and patient record analytics, where variants to develop AI systems serve different populations of patients.
- **Improved Model Generalization and Performance:** Therefore, Federated Learning improves model generalization as a valuable benefit of medical AI. Because FL serves multiple institutions and uses decentralised datasets to train the models, the model is not overtrained in a given dataset. This results in improved results on a large sample size, demographics, disease type and patient conditions, making the model more effective in real-world situations. It also helps to create models for the AI, which would improve its accuracy when dealing with medical conditions and imaging information, thereby coming up with better predictions and diagnoses in practice.
- **Regulatory and Ethical Compliance:** According to different reports compiled by numerous authorities, the healthcare sector is one of the most strictly supervised sectors in the world. It also maintains the validity of the patient's records by setting up stringent guidelines, for example, in the United States: Health Insurance Portability and Accountability Act (HIPAA) or the European Union: General Data Protection Regulation (GDPR). However, due to the Federated Learning approach, the data remains in its original place; hence, it becomes easier to follow these regulations since the raw patient data does not have to travel to different places. This regulatory compliance makes it appealing to healthcare providers interested in implementing AI in their systems without compromising on legal aspects of privacy.
- **Real-world Applications and Advancements:** FL is starting to be used in real-world medical AI challenges. For instance, medical imaging is being used for collaborative training of deep learning models for procedures such as brain tumor segmentation or chest X-ray analysis with no need for transferring personal sensitive patient information to a central location. Federated learning has another potential use in predictive health care, where EHRs located in different institutions have to be used to predict patient outcomes. These examples evidence the application of the Federated Learning paradigm on large-scale use cases and its effectiveness in breaking silos that have been setting back significant advancements in medical Artificial Intelligence.
- **Future Potential and Challenges:** However, there are some limitations to employing Federated Learning in medical AI. Of course, as it has been noted, there are certain challenges within the concept. Factors such as data distribution may be categorized as data heterogeneity; they may be data collected and stored from various institutions and may not be homogenously distributed or labeled in the same manner. Also, acquiring updates of these models and distributing them to clients and servers consumes time, given that large models may be involved. Mitigating these challenges will help achieve the full benefit of federated learning in transforming medical artificial intelligence and enhance the development of privacy-preserving artificial intelligence systems in healthcare. So, in identifying Federated Learning in Medical AI as the main topic of discussion, this paper denotes these insights: Therefore, it is suitable for the healthcare sector because it can protect data privacy when used by large datasets, and further developments in encryption, secure aggregation and improvement of communication protocols would only promote the future capabilities of blockchain solutions.

1.2 Importance in Healthcare

Healthcare data is structurally complex since it is a large collection of information that can be correctly described as heterogeneous; it includes radiology images, EHRs, genomics data, and patient medical history, among others. Every healthcare institution gathers data in its own format that may vary in the set of protocols and standards. This heterogeneity resulted in the inability of the models that would have been developed to generalize well across the population. But, as mentioned earlier, Federated Learning (FL) can solve this problem because learners only train models on local data, and all of them contribute to developing a general model without sharing their databases. [5,6] This makes FL train the model on various data sets of medical conditions, patient data types and disease progression rates, making the model versatile. This is particularly important in medical applications of artificial intelligence, where it is crucial to account for the variance of some features in the population to make accurate predictions of an outcome, such as disease probability, where the early warning signs vary significantly from patient to patient.

However, most health-related data are very sensitive and require much more protection than any other data, and

many laws regulate the way they should be collected and stored, such as the HIPAA in the US and the GDPR in the EU. Interoperability of the raw patient data at the cabinet level between the institutions is quite perilous as it includes cumbersome chances of data leaks and privacy infringement. FL addresses all these issues because no sensitive information is transmitted outside the local institution during the process. Patients' information cannot be disclosed; only the changes in the model are shared in the network, while the models reside in a local database on the personal computer. It also has the advantage of protecting patient data and enabling legal and ethical compliance in organizations. Specific examples of FL include applications that deal with images: radiology and pathology, genetics (genome), and personalized medicine to address patients' behavior and responses to different treatments or prognoses of outbreaks and patients' conditions. Given that FL is capable of combining decentralized, structurally unrelated data and, at the same time, protecting data privacy, FL is set to become a revolutionary enablement technology for the advancement of AI-assisted healthcare.

2. Literature Survey

2.1 Federated Learning Fundamentals

Federated Learning was first popularized through the FedAvg algorithm, which allowed devices or institutions to train a shared model collaboratively without sending raw data to a central server. This innovation laid the foundation for secure, decentralized learning an ideal match for sensitive sectors like healthcare. Over time, several adaptations of this approach emerged: FedProx added a regularization term to stabilize training across non-IID environments, while FedBN used local batch normalization layers to improve consistency in heterogeneous datasets. Scaffold tackled another core issue client drift by applying control variates. These variants have helped FL mature into a more adaptable and reliable framework.

2.2 FL in Medical Imaging

Medical imaging has been one of the earliest and most promising use cases for Federated Learning. In one of the pioneering efforts, researchers successfully trained models to segment brain tumors from MRI scans collected at different hospitals without ever transferring the actual patient images. The results were nearly on par with centralized deep learning systems, demonstrating that high performance doesn't always require centralized data access. These early successes have since inspired extensions into other diagnostic tasks and modalities, confirming FL's value in real-world clinical settings where data sharing is limited by law or patient consent.

2.3 Privacy Techniques in FL

In order to strengthen privacy assurance in FL, various state-of-the-art cryptographic and data privacy-preserving methods have been incorporated. Differential Privacy (DP) is one of the most commonly used techniques where noise is added to gradients or model updates before broadcasting them to the central server. This assists in protecting individual items of data to be inferred even from the aggregated results. Consequently, Homomorphic Encryption (HE) operates differently since it allows computations to be carried out on the encrypted data; thus, the raw data and even further computations are kept secure. Nevertheless, HE contains provisos, which can be computationally expensive. The other is Secure Multiparty Computation (SMC), whereby a number of parties can compute a function of inputs to which no other party gains knowledge of inputs of the other parties. These techniques offer a strong means of protecting privacy in FL settings, especially when applied to sensitive contexts such as health and monetary services provision, but they are often less efficient computationally and in terms of communication.

2.4 Challenges Identified

Unfortunately, several issues with Federated Learning are crucial and currently prevent it from being widely adopted in the industry. The main drawback is, in fact, the communication overhead because FL entails the continual exchange of the model between the clients and the center server. The limitation associated with this is that it can be challenging in large-scale or bandwidth systems environments. Another issue that can be problematic is data heterogeneity, which results when the datasets possessed by different clients are non-IID, resulting in lower accuracy of the resultant model and instability in convergence. Unfortunately, there is no consensus on evaluating, comparing, and, most importantly, rolling out such systems. Finally, FL systems have issues of security risks such as poisoning attacks, which is a process by which a client pollutes the model, and inference attacks, in which a client tries to extract information from the updates. Solving these challenges remains a topic of interest in developing secure, efficient and robust FL systems.

3. Methodology

3.1 System Architecture

The FL system is used to provide the ability to train the model with private data of patients to several medical institutions, for example, hospitals. In this setup, each participating hospital can be considered a client, and each client has a local data source that may have very private medical imaging data such as MRI scans, X-rays or CT scans. [11-15] Instead of uploading the data to another server, this might be a privacy, legal or ethical issue; the data is brought to each client site where the machine learning model is trained locally and independently. After the local training, the hospital does not upload the raw data to the server. Still, it uploads only the model parameters or weight updates of the model to an aggregation server.

The central server is useful because it pulls the model updates from each of the participating clients and combines them using Federated Averaging (FedAvg). The updated global models accumulated by each round are then broadcast back to the clients so they can revise their local models. These steps are repeated through several iterations of the communication rounds before reaching an acceptable performance level of the global model. This approach guarantees that the collected and stored medical data are not

disclosed to third parties HIPAA and GDPR standards. For the purpose of improving communication and the level of its protection, some more subcomponents can be included in the system structure. For example, the model updates can be encrypted through homomorphic encryption or protected through secure aggregation techniques. Update leakage can be restricted using methods that belong to the differential privacy category. Therefore, to handle such variance of data distribution across the various hospitals and to obtain the best results for each specific case, more elaborate techniques like adding new layers for a particular hospital or fine-tuning for the local region may be applied. One particular type of architecture for this is especially suitable for medical research and diagnostic tool creation in large numbers of patients, where data variety and exchange are essential but cannot be centrally stored due to privacy issues. All in all, the architecture of the FL system presents a reliable and efficient solution in the context of privacy-based ML across different institutions in the medical context.

3.2 Training Process

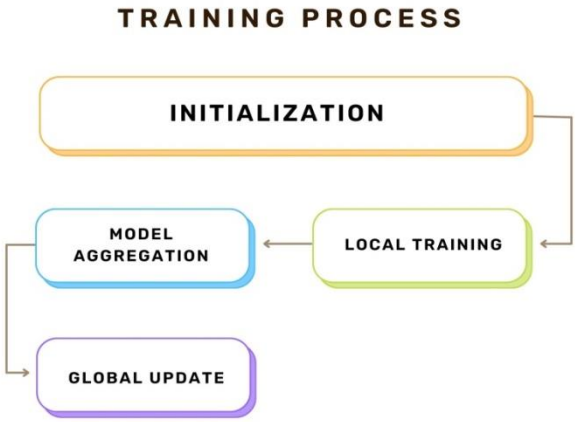

Figure 2: Training Process

- **Initialization:** This training process starts right from the central server of the network by invoking a global Machine Learning model. It has either distributed random or pre-trained weights or is used as the primary model for all clients. This is important mainly because it puts all the clients at the same basic point whereby they are equally informed, which is important in a distributed training environment. It is then transmitted to the individual clients (such as hospitals in the case of medical FL) for the first round of local updates.
- **Local Training:** After collecting the global model, each client tunes this model for local training on their respective local data set. This training is carried out separately and commonly takes place for one or several local epochs. At this stage, clients apply certain modifications to the model weights due to differences in their distribution data, and institutions might have different data distributions. Local training leverages valuable patterns from each client's data set while preserving the patient's data in-house, especially to avoid leaks.
- **Model Aggregation:** Finally, unlike the raw data, each client transmits its weights, obtained after the local update, back to the central server. The server then performs model aggregation, where the most often used algorithms are the Federated Averaging (FedAvg). This implies a procedure of averaging the received model parameters by using a correct weight often arrived at from the size of each client's data. Aggregation allows the construction of a better model from all clients' knowledge, as it may be more diverse due to the distribution of the data.
- **Global Update:** After the aggregation, the global model weights are updated with new average values received from the other parameters. It then forms the new global model shared with all the clients at the start of the new training round. Clients replace the previous model with the updated version, and local training for the new model begins. This iterative process goes on until either the achievement of the global model performance or depending on the set convergence criteria.

3.3 Privacy-Preserving Techniques

- **Differential Privacy:** Differential Privacy (DP) is an approach focused on anonymising individual occurrences used for training the model by adding stochastic noise, normally from Laplace or Gaussian distribution. They specially introduce this noise into the gradients or the model parameters before it is transmitted to the central server in the schema of Federated Learning. This enables minimizing the impact of any given data point; hence, when the model updates are compromised, it will be hard for the adversaries to come to any conclusions. DP has a concrete and numerical privacy measure, which is denoted often with (ε, δ),

and is crucial when the data to be trained is sensitive, such as the medical data.

Figure 3: Privacy-Preserving Techniques

- **Homomorphic Encryption:** Homomorphic Encryption (HE) implies carrying out data computations without decrypting the message while still being secure. This capability is even more valuable in Federated Learning, where the model updates can be encrypted on the client side by, for instance, Paillier or CKKS (Cheon-Kim-Kim-Song). These encrypted updates are then passed to the central server, where some forms of aggregation can be performed without a direct peek at the original data. After aggregation, the final encrypted result can only be unlocked by the correct entity; only the entitled individuals can decrypt it. It maintains data confidentiality for all the data during model training and aggregation but incurs some more computation time.
- **Secure Aggregation:** Secure Aggregation is a method to perform aggregation where the server can add up all the client's model updates but cannot learn updates from any of them individually. Every client sends a change of parameters encrypted with random masks or through cryptographic keys, whereby the server does not realize the exposure of individual parameters. The server may then cancel out the masks (with the help of the clients) so that he does not observe any of the contributions in plaintext as he/she adds the sum of the updates. This technique averts the possibility of the server or any other unauthorized person from helping reconstruct federated data, thus making Federated Learning more secure.

4. Result and discussion

4.1 Experiment Setup

This experiment is planned to create an environment that could approximate the Federation Learning system's nature in which several hospitals (as clients) learn jointly using the central model without sharing patient data. Namely, there are 5 replicas of hospitals used in the computations, and each of them has its medical data, which is necessary for healthcare purposes. Both datasets used in the experiment are NIH Chest X-ray and MIMIC-III datasets, which are quite popular among medical researchers. The chest X-ray dataset from NIH is one of the biggest pools of frontal chest X-ray images, with more than one hundred thousand images with labels for various diseases in lung disorders. The MIMIC-III dataset, on the other hand, is a comprehensive source of clinical data since it offers a variety of data from EHRs of critically ill patients. In this way, the experiment would be able to evaluate the efficiency of Federated Learning for both the medical imaging data and the clinical record information. This setup also allows testing not only the effectiveness of Federated Learning in healthcare but also the potential privacy risks when using such techniques on identifiable health information, which makes such a scenario highly applicable to the actual healthcare practice.

Table 1: Model Performance Comparison

Method	Accuracy (%)	AUC (%)	Privacy Loss (ϵ)
Centralized	94.8%	95%	0%
FL (No DP)	94.2%	94%	0%
FL + DP	92.7%	91%	3.2%

- **Centralized Method:** The Centralized model can also be referred to as the traditional model, as it deals with data collected from all the hospitals where data from one or several hospitals feeds into the model. This yields the highest accuracy of 94.8% and the highest area under the curve of 95%, which is the case since the model can use the entire dataset in training. However, one main disadvantage of this approach is the need to disclose patient data that may contain private information, which is easily breached in the healthcare environment.
- **Fl without Differential Privacy (also described as FL-No DP):** The FL (No DP) model can be considered Federated Learning, where no privacy-preserving measures were applied. Here, you train the model locally for each hospital, and only the model weights are transferred to the central server, not the raw data itself. Therefore, the accuracy achieved in this case is 94.2%, and the AUC of 94 % is slightly lower than that of the centralized model. This paper reduces such risks by not sharing

sensitive data but does not eliminate them entirely since the model updates could be attacked.

- **The integration of federated learning with differential privacy is examined as [FL + DP]:** As for the second aspect in FL + DP, DP reconfigured the model updates so that it would not leak any information about the patients. By incorporating DP, noise is incorporated into the model updates, leading to a privacy loss of 3.2%. This is at the expense of the model's performance since the accuracy reduces to 92.7%, and the AUC reduces to 91%. Nevertheless, this leads to performance degradation since the DP technique offers greater privacy protection; thus, it is preferable if privacy is a major concern. These results show that the model's usefulness is an exchange for protection, which the federated learning accompanied by \DP provides rather adequately.

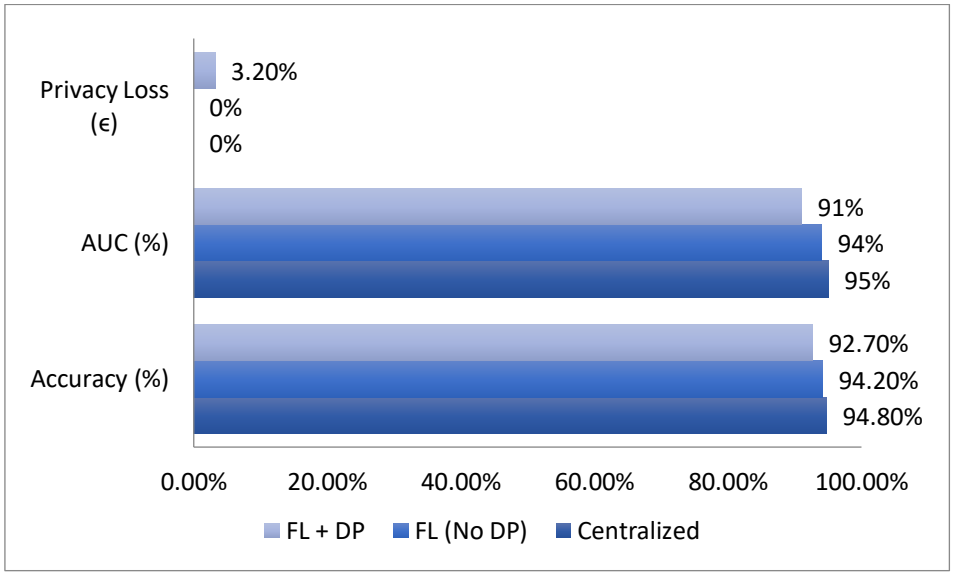

Figure 4: Graph representing Model Performance Comparison

4.2 Findings

- **Minimal Drop in Accuracy with Privacy-Enhancing Techniques:** The application of the DP to FL models led to a decrement in accuracy and the AUC; they reduced from 94.2% to 92.7%, respectively. This decline, however, was relatively moderate and, thus, shows that integrating approaches such as DP does not necessarily decrease the model's performance significantly. The reason for a slight decrease in the model's efficiency is the noise introduced to the updates to prevent information leakage from individual data samples. The noise must make the information difficult to be learned by the model, hence its effectiveness in preventing the extraction of information from the model. However, it slightly hinders the learning process, leading to a small margin of decrease in model accuracy and AUC. However, the small amount of sacrifice indicates that DP can indeed offer privacy protection while not greatly affecting the model's performance; therefore, it is fit for use in scenarios where data ownership must be kept private.
- **FL Outperformed Siloed Learning Models:** The statistics presented revealed that FL's performance was always higher than the isolated local training where model training occurs independently without any FL interaction. Consequently, in this experiment, it was possible to record an improvement ranging from 8-10 per cent compared to standalone models. This can be credited to FL's feature of working in a decentralized fashion where multiple hospitals train one model but use diverse datasets without data sharing. Training on more varieties of data sources makes the model more generalized so that it can handle wide variations of various medical conditions as may be exhibited in different hospitals. This result further establishes the flexibility and efficiency of FL in deriving better models from distributed data with confirmation of privacy and data protection without compromising the signal quality and integrity and out-competing the centralized approach.
- **Communication Efficiency Improved with FedAvg + Compression Algorithms:** When it comes to large models and many clients, communication overhead is a significant constraint in Federated Learning. However, the integration of FedAvg with model compression algorithms helped limit the amount of data to be exchanged between the clients and the central server to a great extent. Optimizers reduce the size of the updated model since the amount of communication in each round and, hence, the network bandwidth is limited. This improvement is important in real-world scenarios since the cost of data transfer and the use of communication channels may be significant. As the

frequency of updates is less, the system becomes less resource-intensive. As such, Federated Learning becomes more scalable and practical in scenarios where the network is not as strong as in rural hospitals or where the latency is high.

4.3 Discussion

- **Trade-off Between Privacy and Model Performance:** The results reveal that with the use of DP, it is possible to have a trade-off between privacy and the model's performance. Although DP is a useful method of preventing the leakage of sensitive information in the model updates, it decreases the model's accuracy due to the noise added to the updates. When applying DP, both the accuracy and the AUC of the model were lower than before, which is expected due to the introduction of noise in the data learning process. However, this is mainly true with the privacy hyperparameter, which is the major factor determining the DP's feasibility in different practical applications. A smaller value of ϵ introduces stronger privacy in the sense that fewer characteristics of the individual data points can be learned from the construction of the model, but it also raises the noise level and, thus, a higher loss in performance of the model.
- On the other hand, a larger value of ϵ means lesser noise that improves the model performance. On the other end, the user's privacy is at risk. Thus, there is a need to identify an optimal choice of ϵ such that sufficient levels of privacy are achieved at the same time the performance penalty incurred is kept relatively small.
- **Secure Aggregation Enhances Trustworthiness in Real-World Settings:** Secure aggregation plays a huge role in providing a trustworthy environment in Federated Learning, mainly when implemented in real-life applications such as healthcare. The type of communication used is that the central server collects model updates from clients without accessing the individual updates, yet maintaining the confidentiality of data. This eliminates the possibility of disclosure of sensitive information in the process of model amalgamation, which raises insecurity issues. When it comes to healthcare organizations that may deal with patient information, aggregation must be more secure. It also makes sure that even the server or any other middleman participant cannot have the ability to put together or guess at patient-specific details from these updates, which would help make all participants in the federated learning system put their trust in the arrangement. Such trust is necessary for adopting Federated Learning in areas or regions where privacy is crucial, like hospitals or clinics.
- **Differentially Private Models Require Tuning of Privacy Budget (ϵ):** Although DP has proven to be one of the most effective ways of preserving data privacy, using the privacy budget (ϵ) is critical in determining its efficiency. A very small ϵ value distorts the updates of the model to a large extent, which in turn results in the handicaps of the ability of the model to learn good features from the data, thereby causing significant performance degradation. While a high ϵ value has the advantage of increasing noise to reduce noise that affects the outcome of the learned model, it decreases the privacy level. Therefore, attackers have a high chance of learning some important customer information. Thus, the challenge ensues on how much noise is optimally introduced into the model and the usefulness of the model that ensues from the noise. In order to achieve the above objectives, there is a need to fine-tune and develop a deeper understanding of the model so as not to give out false results while keeping it secure. Moreover, the value of ϵ is also sensitive to the application type; that is, some applications are more sensitive to privacy as well as optimizing the trade-off between privacy and other factors.

5. Conclusion

FL is a revolutionary approach to cooperative healthcare research as it provides a way for training the ML algorithms across various institutions with the help of patients' data protection. In this regard, FL empowers several hospitals or other medical centers to contribute to creating the model while medical data remains decentralized yet secure owing to DP and secure aggregation. Our work also demonstrates the importance of using FL to increase diagnostic accuracy because models that learn from data from other institutions are more generalized and less sensitive to variations in patient data, which is important in healthcare settings. Also, FL can support the implementation of AI in different institutions since the model updating happens with the participants' validation without sending the actual patient data, which remains a significant concern in the healthcare industry due to security issues. Although the values of FL are rather high, several obstacles have to be overcome. This is due to challenges such as high communication costs and varieties of data distribution across institutions, which are referred to as heterogeneity.

However, the study shows that these issues are constantly being worked on and solved by advancing encryption techniques and secure protocols, thus making implementing FL more viable. As for future perspectives, the development of Federated Learning in healthcare is expected to proceed in the direction of defining proper FL pipelines that push the technology's deployment along different institutions and districts. One such extension is to combine FL with blockchain, which will ensure the accountability of the model development process. Moreover, FL's continued development could make its integration into real-time clinical decision support systems, which would be a critical application of AI in patient care in conjunction with data protection. In conclusion, I have shown that Federated Learning is the promising approach to support the privacy-

preserving development of AI in healthcare and give medical researchers and clinicians eager for collaborative and data-driven progress in treating and combating diseases an opportunity to make positive changes.

References

[1] Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2020). Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems, 2, 429-450.

[2] Li, X., Jiang, M., Zhang, X., Kamp, M., & Dou, Q. (2021). Fedbn: Federated learning on non-iid features via local batch normalization. arXiv preprint arXiv:2102.07623.

[3] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S., Stich, S., & Suresh, A. T. (2020, November). Scaffold: Stochastic controlled averaging for federated learning. In International conference on machine learning (pp. 5132-5143). PMLR.

[4] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (pp. 1175-1191).

[5] Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019, May). Exploiting unintended feature leakage in collaborative learning. In 2019 IEEE symposium on security and privacy (SP) (pp. 691-706). IEEE.

[6] Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. Scientific reports, 10(1), 12598.

[7] McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017, April). Communication-efficient learning of deep networks from decentralized data. In Artificial intelligence and statistics (pp. 1273-1282). PMLR.

[8] Rieke, N., Hancox, J., Li, W., Milletari, F., Roth, H. R., Albarqouni, S., ... & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ digital medicine, 3(1), 119.

[9] Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 1-19.

[10] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (pp. 308-318).

[11] Gentry, C. (2009, May). Fully homomorphic encryption using ideal lattices. In Proceedings of the forty-first annual ACM symposium on Theory of computing (pp. 169-178).

[12] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), 1-210.

[13] Li, Q., He, B., & Song, D. (2020). Practical one-shot federated learning for cross-silo setting. arXiv preprint arXiv:2010.01017.

[14] Nasr, M., Shokri, R., & Houmansadr, A. (2019, May). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In 2019 IEEE symposium on security and privacy (SP) (pp. 739-753). IEEE.

[15] Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with non-iid data. arXiv preprint arXiv:1806.00582.

[16] Loftus, T. J., Ruppert, M. M., Shickel, B., Ozrazgat-Baslanti, T., Balch, J. A., Efron, P. A., ... & Bihorac, A. (2022). Federated learning for preserving data privacy in collaborative healthcare research. Digital Health, 8, 20552076221134455.

[17] Haripriya, R., Khare, N., & Pandey, M. (2025). Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings. Scientific Reports, 15(1), 12482.

[18] Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., ... & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. ACM Computing Surveys (Csur), 55(3), 1-37.

[19] Jin, Y., Zhu, H., Xu, J., & Chen, Y. (2023). Federated Learning. Springer Nature Singapore, Singapore.

[20] Naithani, K., Raiwani, Y. P., Tiwari, S., & Chauhan, A. S. (2024). Artificial Intelligence Techniques Based on Federated Learning in Smart Healthcare. In Federated Learning for Smart Communication Using IoT Application (pp. 81-108). Chapman and Hall/CRC.

[21] Bashir, A. K., Victor, N., Bhattacharya, S., Huynh-The, T., Chengoden, R., Yenduri, G., ... & Liyanage, M. (2023). Federated learning for the healthcare metaverse: Concepts, applications, challenges, and future directions. IEEE Internet of Things Journal, 10(24), 21873-21891.