ISSN: 2959-6386 (Online), Volume 3, Issue 3, Sep 2024

Journal of Knowledge Learning and Science Technology

Journal homepage: https://jklst.org/index.php/home

JKLST
INNOVATIONS IN INTERDISCIPLINARY RESEARCH AND PRACTICES FOR A SUSTAINABLE FUTURE

Comprehensive Review

DEVELOPING PRIVACY-PRESERVING FEDERATED LEARNING MODELS FOR COLLABORATIVE HEALTH DATA ANALYSIS ACROSS MULTIPLE INSTITUTIONS WITHOUT COMPROMISING DATA SECURITY

Basirat Oyekan

Lamar University, Beaumont, Texas, USA

Abstract

Federated learning is an emerging distributed machine learning technique that enables collaborative training of models among devices and servers without exchanging private data. However, several privacy and security risks associated with federated learning need to be addressed for safe adoption. This review provides a comprehensive analysis of the key threats in federated learning and the mitigation strategies used to overcome these threats. Some of the major threats identified include model inversion, membership inference, data attribute inference and model extraction attacks. Model inversion aims to predict the raw data values from the model parameters, which can breach participant privacy. The membership inference determines whether a data sample was used to train the model. Data attribute inference discovers private attributes such as age and gender from the model, whereas model extraction steals intellectual property by reconstructing the global model from participant updates. The review then discusses various mitigation strategies proposed for these threats. Controlled-use protections such as secure multiparty computation, homomorphic encryption and conidential computing enable privacy-preserving computations on encrypted data without decryption. Differential privacy adds noise to query responses to limit privacy breaches from aggregate statistics. Privacy-aware objectives modify the loss function to learn representations that protect privacy. Information obfuscation strategies hide inferences about training data.

Keywords: Federated learning, healthcare, security aggregation, homomorphic encryption, membership inference, model inversion, collaborative learning, medical imaging, genomics, precision medicine, data heterogeneity, computational efficiency, explainable AI

Article Information:
Received: 12-June-24 *Accepted:* 27-July-24 *Online:* 25-Aug-24 *Published:* 25-Sep-24

DOI: https://doi.org/10.60087/jklst.vol3.n3.p139-164

[i] **Correspondence author:** Basirat Oyekan **Email:** basiratoyekan2019@gmail.com

1. INTRODUCTION

1.1. Emergence of FederatedLearning

Federated learning is an emerging distributed machine learning technique that enables collaborative training of models among devices and servers without exchanging private data (McMahan et al., 2017). In conventional centralized machine learning, all the training data need to be sent to a central server for model training, which poses serious privacy and security risks because of the sheer volume of sensitive personal and health data involved. To address this issue, federated learning allows the training process to occur locally on personal devices such as mobile phones and Edge/Cloud servers, after which only the model parameters are exchanged without revealing the underlying private data. This promotes collaborative learning across institutions while protecting individual privacy (Li et al., 2018). Over the past few years, federated learning has gained significant attention from both industry and academia for its ability to learn from decentralized data in a privacy-preserving manner (Google AI Blog, 2017).

Owing to its decentralized and privacy-preserving nature, federated learning is highly suitable for healthcare applications involving collaboration across multiple institutions for tasks such as disease prediction, diagnosis, and treatment optimization. Some key works include predictive modeling of clinical outcomes for COVID-19 patients through federated learning across different hospitals (Dayan et al., 2021). Another study performed end-to-end private model training across diverse medical imaging datasets from multiple centers for tasks such as segmentation (Kaissis et al., 2021). Federated averaging enables the distributed training of models for automated diagnostic analysis of medical images without the need to share patient data (McMahan et al., 2017).

Nevertheless, federated learning is not safe from privacy and security threats, which must be resolved for effective and safe application in sensitive fields such as healthcare. Model parameters and updates can be used to leak privacy within model inversion, membership inference, and attribute inference attacks. Moreover, it is also possible to steal the intellectual property of locally trained models. In addition, there are more technical problems associated with data heterogeneity and distribution and differences in the structure of the institutions and organizations that participate in the collaboration. Furthermore, there is a need for strong methods to achieve privacy, utility and inefficiency for the complete implementation of federated learning.

1.2. Privacy and Security Risks Federated Learning

Unlike federated learning, which decentralizes data, model updates exchanged from devices to servers during training can leak sensitive information (Shokri and Shmatikov, 2015; Melis et al., 2019). They mentioned that the adversaries can use them to learn some of the parameters of the training data, such as membership, properties or reconstructed raw values via model inversion, attribute inference or model extraction attacks (Nasr et al., 2019; Song et al., 2017; Fredrikson et al., 2015). Furthermore, performing distributed training in multiple institutions comes with adversarial drill risks of attackers injecting poisoned updates or attackers deducing the propriety characteristics of other participants' local models (Hitaj et al., 2017; Bagdasaryan et al., 2020). Thus, even if raw data are not shared, parties still want assurances that their private training data and local model intelligence are protected during the collaborative learning process (Fletcher and Islam, 2020). This hinders wider adoption of federated learning for applications involving sensitive data domains such as healthcare.

1.3. Overcoming Privacy and Security Risks

Several privacy-preserving techniques have been proposed and analyzed to counter privacy threats in federated learning and enable its secure use on sensitive health data

(Jayaraman and Evans, 2019; Kaissis et al., 2021). Differential privacy is a framework that can limit privacy breaches from model updates by adding calibrated noise (Dwork, 2006). Homomorphic encryption enables computations directly on encrypted model updates for outsourced computations (Gentry, 2009). Securing multiparty computations distributively trains a model among parties without directly sharing individual updates (Bonawitz et al., 2017). Information obfuscation hides informative patterns while preserving utility via techniques such as vector perturbations (Wang et al., 2020) or generative models (Hitaj et al., 2017). Privacy-aware objectives modify traditional learning objectives to learn privacy-respecting representations (Phong et al., 2018).

However, balancing privacy, utility and efficiency remains challenging. Furthermore, comparisons of such approaches against different types of attacks are still lacking because of their reliable use in healthcare applications (Zhang and Zhu, 2021). This survey aims to provide a systematic overview of key privacy and security risks in federated learning and analyzes existing mitigation strategies to realize their full potential.

1.4. Research gaps

While significant advances have been made, some important research gaps remain in the development of robust privacy-preserving federated learning models for practical collaborative health data analysis across institutions:

- Comprehensive benchmarking and evaluation: Standard frameworks and metrics are needed to benchmark privacy, utility and inefficiency of techniques against different threats (Karargyris et al., 2021; Melis et al., 2021). Real-world datasets and models are also crucial.
- Balancing privacy-utility-efficiency tradeoffs: Achieving an optimal balance remains challenging because of modeling assumptions and implementation overheads (Chen et al., 2020; Hitaj et al., 2017). Adaptive techniques that match privacy levels to data sensitivity are needed.
- Addressing multiple threats: Existing works address mainly specific attacks in isolation. Combinations of defenses that handle a broad threat surface are lacking (Nasr et al., 2019; Wang et al., 2019).
- Scaling to complex heterogeneous healthcare data: Most works consider simple tasks or assume IID data, unlike real healthcare applications involving diverse, skewed multiinstitutional data (Kaissis et al., 2021; Sattler et al., 2020).
- Bridging theory and practice: Practical considerations in implementing robust, scalable and compatible privacy techniques in production systems are still underexplored (Melis et al., 2019; Balle et al., 2021).
- Standards and regulations: A lack of universal standards and unclear regulations hinder widespread ethical use of privacy techniques in healthcare (Christen et al., 2021; VOigt and Bussche, 2017).

This review aims to address these gaps by providing a comprehensive analysis of threats and mitigation strategies in federated learning.

1.5. Purpose of The Review

The main purpose of this article is to provide a systematic review of key privacy threats in federated learning and analyze existing mitigation strategies to enable trustworthy collaborative health data analysis across multiple institutions without compromising user privacy and data security. It discusses open challenges and outlines future directions for developing robust privacy-preserving federated learning models suitable for practical healthcare applications involving sensitive patient data.

1.6. Review Aims and Objectives

The aims of this review are as follows:

1) Initially, it is important to provide a broad picture of the numerous privacy risks that will be associated with waiting for the federated learning models when they are employed for collaborative analytics of health information across institutions. This involves discussing attacks that target leaked information from model updates such as membership inference, attribute inference and model extraction.

2) To analyze the state-of-the-art techniques proposed to counter these privacy threats and enhance security in federated learning. This covers defenses under differential privacy, secure computation, information obfuscation and privacy-aware learning paradigms. The objective is to understand how existing methods address specific threats.

3) To evaluate the practical considerations and limitations of existing privacy-preserving techniques when they are applied to real-world healthcare use cases. This involves analyzing the trade-offs between privacy, utility and efficiency of different methods.

4) To identify important open challenges and research gaps in the development of robust privacy techniques that can balance privacy, utility and scalability for complex healthcare applications involving heterogeneous multi-institutional data.

5) To identify recommendations and prospective studies of standardized evaluation and adaptive techniques, multiple threats should be recognized, and the integration of theory and practice to achieve a reliable federated learning system for collaborative health analytics.

The general objective of this review is to offer guidelines for the safe implementation of federated learning in sensitive areas such as healthcare.

2. LITERATURE STUDY OF PRIVACY-PRESERVING FEDERATED LEARNING IN HEALTHCARE

2.1. Emergence and Applications of Federated Learning in Healthcare

2.1.1. Evolution of Federated Learning Techniques

Federated learning then emerged as a machine learning approach aimed at training models in devices and servers in parallel without sharing users' data. As referred to by McMahan et al. (2017), FL frameworks loop through local client-sided training on the supplied training data and then aggregation of the distributed local computed model updates. It also helps consolidate disparate datasets in various institutions to build models that are much more general with far higher throughputs (Dayan et al., 2021). In the words of Li et al. (2018), federated learning allows model training on a distributed set of data, but no actual data are shared. The authors note that this is done in a cyclical manner where local models of on-user devices are trained and the updates of the parameters are collected by a master server for the creation of the global model. They proposed a novel global model that is then sent back to the devices for additional local training, which makes the global training of the models possible without sharing the actual data of the users.

Analyzing the potential of federated learning in the healthcare domain, Dayan et al. (2021) reported that it can be effective. This study established clinical outcomes for COVID-19 patients by training models on hospital patient data with the ability to share information. This approach relies on the distributed form of medical data while maintaining the privacy of the patient's information, which increases the performance of the models compared with the models trained on the enclosed databases. Two other areas that have been discussed earlier as promising applications of federated learning in healthcare include automated diagnostic analysis of medical images on the basis of distributed model training across datasets originating from different medical centers (McMahan and McMahan 2017). Through decentralized training of models using different imaging data without sharing data with a central repository, federated learning provides a means of creating accurate diagnostic models that are generalizable without compromising patients' confidence.

2.1.2. Applications in Predictive Healthcare Modeling

Initial use cases of federated learning in healthcare were therefore aimed at building risk assessment models for COVID-19-affected patients. Dayan et al. (2021) reported that federated learning allows the training of models on data from multiple hospitals without raw data exchange due to its sensitivity. This approach capitalizes on the spread of medical data while keeping patient data private, which helps enhance the performance of the model compared with models that are based on closed data. It has also been used in a federated learning manner for automated diagnostic analysis of medical images. In 2017, McMahan et al. reported that federated learning methodologies were utilized to train models across different datasets with origins in different medical centers. This approach helps build models with more accurate diagnoses and is more generally applicable without the need to centralize patients' data, which may lead to data leakage.

In their study, AlBadawy et al. (2018) described the approach of federated learning to perform segmentation of brain tumors across multiple different institutions without exchanging patient data. This approach utilized the distributed structure of medical imaging data, but the patients' data were still preserved, which led to better model performance than models trained with centralized medical imaging data.

In addition to predictive modeling and medical image analysis, federated learning has also been applied for the use of precision medicine involving collaborative analytics on multiomics and multimodal clinical records in Liang et al., 2015; Topol 2019. Through disjointed healthcare data scattered across various institutions, federated learning can support the creation of a more accurate predictive model that can be tailored to specific patients and hence begin personalized medicine.

1.3. Multi-institutional Medical Imaging Analysis

One of the most significant use cases of federated learning in healthcare is the E2EE PMT for various medical imaging datasets of multiple institutions for tasks such as segmentation.

As shown by Kaissis et al. (2021), it enables state-of-the-art results, and it is fully compliant with strict patient privacy constraints because of the lack of data sharing.

It has also been applied in the federated learning scenario for multiple medical site collaborations for predictive modeling of chest X-rays. Necessary innovations were described by Dunnmon et al. (2019), who introduced a federated learning method that allowed the training of models on chest X-ray images collected at 13 medical sites and that shared no patient information. This approach capitalizes on the distributed nature of medical imaging data while maximizing the security of the data, and the model performs better.

Albadawy et al. (2018) reported that federated learning was useful for the segmentation of brain tumors, which was performed across different institutions without sharing patient data. Owing to the ability to maintain data privacy in collaborative training of distributed medical imaging data, federated learning has the potential to enhance numerous healthcare-related tasks, such as computer-aided diagnosis and disease monitoring. This makes federated learning challenging for multi-institutional medical imaging analysis, which, in turn, shows its potential for application in healthcare regions. According to Kaissis et al. (2021), federated learning allows the analysis of different medical datasets from various institutions without compromising patients' privacy, thus opening the way to the construction of more accurate and less biased models for numerous clinical uses.

1.4. Enabling Precision Medicine through Collaborative Analytics

Using the available uncoordinated and decentralized health data, federated learning can support precision medicine activities involving the analysis of multiomics and multimodal patient records (Liang et al., 2015). As emphasized by Topol, this approach allows the creation of more precise models for prognosis that are designed for a particular patient, which is a key component of precision medicine.

When further integrating them with blockchain and edge/cloud computing environments, as shown by Pham et al. (2021), federated learning facilitates new collaborative analytics in digital ecosystems. It applies to numerous applications, including CAD applications and patient monitoring, on the basis of networked medical devices and IoT technologies. Similarly, Haidar & Kumar (2021) highlighted how federated learning holds the promise of facilitating joint healthcare analytics with the help of both edge and cloud computing interfaces. The approach is based on the diverse nature of medical data, which ensures privacy while being effective at being used in different fields, such as the remote control of patients and individual recommendations for therapy.

The adoption of federated learning together with blockchain and edge/cloud technologies is likely to transform healthcare provision. In this way, through providing the ability to train a model on a number of data sources at once, to protect data privacy, federated learning can help create unique approaches that focus on patient characteristics and clinical history and effective approaches for precision medicine.

2.2. Privacy Risks in Federated Learning Systems

2.2.1. Membership Inference and Attribute Inference Attacks

In a way, membership updates that are communicated during federated learning training can pose threats of leakage through membership inference, attribute inference or model inversion. Shokri & Shmatikov (2015) noted that malicious parties can tell whether a given data record was used during training or not to guess sensitive aspects of the client, such as their age and gender, among others, just by the parameters of the model. Membership inference attacks on federated learning systems were described, and the potential privacy risks were discussed by Nasr et al. (2019). Their work showed that an adversary is able to see if a particular data record was used to train the global model if a party disclosed updated parameters, which was an invasion of an individual's privacy.

Another actual threat to privacy in the context of federated learning is attribute inference attacks. Shokri & Shmatikov (2015) demonstrated that black-box evasion attacks can reveal sensitive attributes of training data even if they are identical, such as patients' conditions or demographic details. These private attacks require strong defenses in federated learning, especially in sensitive areas such as health care. Appropriate measures that should be adopted to seal gaps that enable the leakage of patients' private data remain key factors in promoting the ethical use of federated learning in medicine.

2.2.2. Model Inversion to Reconstruct Training Data

Model inversion is also used to reverse engineer the raw data back from the model and, as such, is intrusive to participant privacy. Using an example, Fredrikson et al. (2015) showed that one can produce the original training data used in creating a model through probing by asking questions such as 'what do you know about this picture', and hence, despite implementing differential privacy, the model will be at risk of attack. Model inversion attacks were studied by Carlini et al. (2018) in the context of differentially private deep learning models. In their evaluation, they learned that an attacker could reverse engineering and gain identifiable training data samples with ease from the model parameters, much to the chagrin of differential privacy solutions.

Fredrikson et al. (2015) noted that other model inversion attacks are serious and illustrated that even models that were developed with anonymized data are susceptible to the infringement of privacy. Their work focused on how to protect such systems from threats, such as reconstruction threats, that can distort training data on federated learning programs. Given the increasing use of federated learning in healthcare applications, mitigating these risks is highly important for protecting patients' identities. This means defending against the potential reconstruction of sensitive medical data from collaboratively trained models to check the excellent and trustworthy application of federated learning methods in the clinical domain.

2.2.3. Model Poisoning and Property Theft Attacks

Attackers can perform poisoning to control updates of the global model or even gain intellectual property information from other participants in federated learning systems. In his 2017 work, Hitaj et al. explored the effectiveness of a model poisoning attack in which adversaries can control and influence the global model by feeding lethal local variations during training. Bagdasaryan et al. (2020) investigated how federated learning systems are exposed to property theft attacks. They reported that, from the aggregated global model, their adversary was able to obtain information such as the model architecture or the hyperparameters, which is a clear and severe violation of intellectual property rights.

In their paper, Orekondy et al. (2019) focused on model extraction attacks in a federated learning context. Their work showed that it was possible to derive locally trained models for participants by using the parameter updates sent during collaborative training, thus increasing privacy and security risks to individual models. There are some concerns because, in federated learning systems, the participants are spread out across various locations, meaning that various threats and attacks can be executed, such as tampering with the global model or stealing intellectual property. Such proposals require bolstered defense mechanisms and safe procedures for aggregating parameters to prevent them from adversely affecting the integrity and privacy of collaborative model training in federated learning networks.

2.2.4. Increased Risks due to Distributed Training

Owing to the distribution of FL across different systems, the heterogeneity of the systems increases the threat of privacy attacks compared with single centralized learning. As pointed out by Melis et al. (2019), in federated systems, there are no centralized datasets; hence, auditing and, thus, private disclosure are complicated. Carlini et al. (2018) and Song et al. (2017) looked at the potential of collaborative deep learning models for remembering other information, which is potentially a problem from a private point of view, as users are proiled. In their studies, they showed that federated learning models can actually transmit information from training data and are thus a threat to privacy.

That is why Wei et al. (2020) focused on the study of privacy issues in distributed training in federated learning systems. Their work pointed out the lack of privacy considering the new structure of decentralized training and the susceptibility to various attacks from the intruding participants aiming at stealing the data or falsifying the model updates. An increasing number of people and organizations are embracing federated learning, especially in health care and sensitive areas; hence, addressing the increased risks of privacy that come with distributed training is important. Stringent assessment of defense mechanisms, secure data aggregation procedures, and proper auditing mechanisms are needed to protect the privacy of federated learning systems in medical applications.

2.3. Evaluating Defenses against Privacy Attacks

2.3.1. Differential Privacy Framework

Differential privacy is a provable solution that keeps privacy leakage from model updates up to the level in which calibrated noise is added to the query responses. In the case of applying differential privacy to federated stochastic gradient descent, McMahan et al. (2018) reported that the procedure of learning does not allow the analyst to infer with certainty whether a specific sample is part of the training dataset, for which privacy considerations are relatively robust. The theoretical framework for differential privacy was developed in the work of Dwork (2006), and it is now used as a key method to protect individuals' privacy while performing data analysis and machine learning on big data. What differential privacy does involve injecting a computed amount of noise into the computation process by which one can guarantee that the outcome of a query or analysis is almost oblivious to whether a certain record is included in a database.

Previous studies have also investigated the use of differential privacy with respect to the federated learning process to address privacy concerns arising from joint model training. In McMahan et al. (2018), the authors established that the differential privacy added to the federated stochastic gradient descent algorithm stops membership inference attacks, whereby the adversary attempts to establish whether a specific data instance was used in the creation of the global model. Even though differential privacy guarantees privacy, the problem of achieving these theoretical privacy levels when adopted to federated learning systems is that privacy is always lost at the expense of model utility. However, current active research aims to improve these trade-offs and build better techniques that can protect privacy up to the maximum level while incurring a small model quality loss as much as possible in different application areas, such as healthcare.

2.3.2. Homomorphic Encryption and Secure Computation

Homomorphic encryption allows for computation of the encrypted data directly, which also makes federated training outsourceable without decryption. The idea of fully homomorphic encryption that allows unrestricted computations to be performed on an encrypted dataset without decrypting it was initially proposed by Gentry in 2009. Cao et al. (2013) researched how homomorphic encryption works in cases of federated learning. In their study, they showed that it is possible to train machine learning models using encrypted data supplied by multiple parties to the server, and in the process, the parties do not divulge their information to the server or any other party.

Other methods that are based on secure multiparty computations have also been considered in the field of federated learning, for example, secure aggregation. Bonawitz et al. (2017) conducted a study, which led to a safe aggregation approach that allowed parties to collectively train their respective models without necessarily sharing their separate updates, hence enhancing privacy, but at the same time promoting proficiency in distributed training. Although the techniques of homomorphic encryption and secure computation provide reasonable levels of privacy, the actual computations involved are complex, and the cost of communication is high. There is still more work to be done to design and invent more efficient algorithms and systems that can be easily implemented in actual federated learning systems, especially in scarce resource contexts such as healthcare IoT devices.

2.3.3. Evaluating Privacy-Utility Trade-offs

An important aspect of privacy preservation and evaluation of the inefficiency of privacy-preserving methods in federated learning is the analysis of the cost and benefit of using privacy-preserving methods for federated learning. Some of the suggested rigorous simulation settings and benchmarks, such as Federated Learning Datasets (FLData), aimed at assessing the accuracy–privacy cost of different defenses to such attacks as membership inference, were suggested by Melis et al. (2019). Jayaraman & Evans (2019), the author that that we the author revisits a number of decisions on privacy vs utility of federated learning and differential privacy. Their work involved mining how various forms of noise and the amount of privacy introduced would affect the performance and privacy of the models, which is useful when trying to make compromises between the two.

Thus, more recent work by Karargyris et al. (2021) offered a review of privacy preservation methods, with an emphasis on medical imaging programs. Their work compared, inter alia, the effectiveness of differential privacy and secure aggregation when applied to federated medical image learning tasks and an appreciation of the problems posed by the need to protect patient privacy while ensuring model accuracy.

Although a good deal of work has been done on evaluating PUM via simulations and using benchmark datasets, there is still a dearth of real-life membership inference attacks due to data accessibility issues, especially in healthcare environments. Future studies should focus on elaborate assessment models and procedures that utilize diverse actual-world datasets to determine the practical applicability of preservation methods in fed learning systems.

2.4. Privacy-Preserving Federated Learning for Collaborative Health Data Analysis

2.4.1. Federated Learning and Differential Privacy Techniques

Federated learning allows for cooperative model updates across several institutions while maintaining the data at those institutions, which is good for medical big data. As depicted in

Figure 1 shows that federated learning comprises a global model aggregated from several local models that are linked with standardized health record data and special healthcare centers (Loftus et al., 2022). However, recent research has shown that even federated learning models are no longer secure from attacks such as membership inference attacks or unintended memorization of training sets, as identified by Hu et al. (2022) and Carlini et al. (2018). There is an exciting method, differential privacy, which, by adding carefully calibrated noise to the model updates or to the results (Demelius et al., 2023), minimizes such threats to privacy. Adnan et al. (2022a) developed a federated learning framework with differential privacy for segmentation of brain tumors, which proves that privacy and a reasonable level of model performance can be preserved.

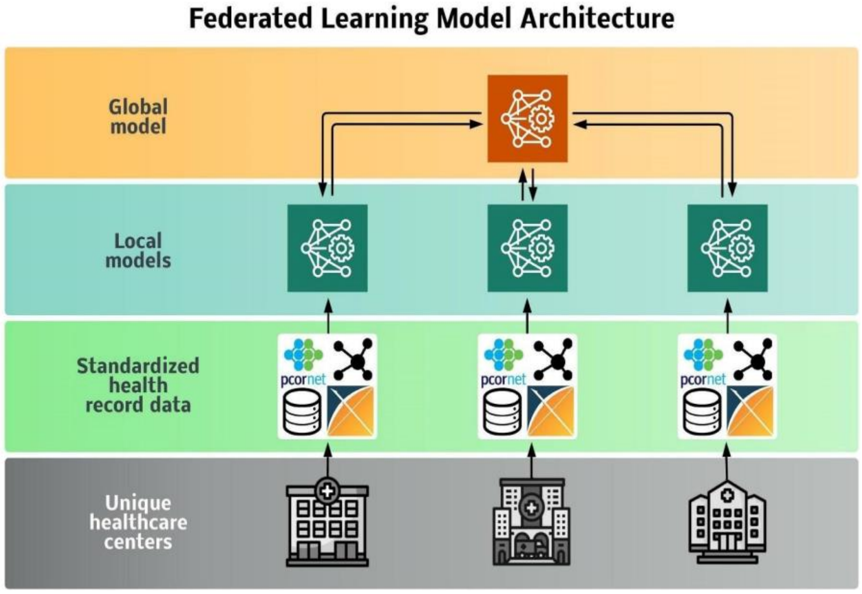
Federated Learning Model Architecture
Global model
Local models
Standardized health record data
Unique healthcare centers
pcornet
pcornet
pcornet

Figure 1. *Federated learning model architecture. Source: Loftus et al. (2022)*

The following figure presents the federated learning model architecture, which is a representation of the architecture of the system used in federated learning. It is a single global model linked with more global models, as many local models exist, accompanied by specific health record data and successive healthcare centers. The different components are identified by different color codes for better understanding, and there are arrows to show that information decreases from one component to the other. Source: Loftus et al., (2022)

Although differential privacy is beneficial in the context of privacy, it reduces the quality of the developed machine learning models, particularly in the application of high-dimensional data, such as medical images (Jarin and Eshete, 2022; Jayaraman and Evans 2019). Some recent works have investigated new ways to increase the utility of the FL approach, which uses secure aggregation protocols (Bonawitz et al., 2017; Huang et al., 2022) or employs trusted execution environments (Ekberg et al., 2014; Buyukates et al., 2022). Kaissis et al. (2021) federated learning to resolve multi-institutional medical image analysis issues while simultaneously utilizing differential privacy and secure aggregation to increase the privacy of the pipeline.

Another approach to privacy-preserving collaborative learning is homomorphic encryption, which allows computations on encrypted data without decryption (Acar et al., 2018; Cao et al., 2013; Gentry and Halevi, 2011). Nevertheless, the incumbrance of high computational complexity is seen as a major hindrance for the adoption of extant homomorphic encryption solutions in healthcare (Kalapaaking et al., 2022). There are currently propositions about using homomorphic encryption in conjunction with either federated learning or differential privacy to potentially open new opportunities for secure and privacy-preserving analysis of medical data.

2.4.2. Federated Learning for COVID-19 Outcome Prediction

COVID-19 has emphasized how many medical centers require joint analysis of large datasets accumulated in various institutions. Dayan et al. (2021) discussed how federated learning is effective in predicting the clinical outcome of COVID-19 patients via data from different healthcare settings. Their proposed federated learning model provided similar accuracy to traditional centralized models of analysis while protecting individual information and addressing parameters such as HIPAA.

In addition to COVID-19, federated learning has been presented for a wide variety of other medical applications: chest radiography imaging classification (Dunnmon et al. in press), brain tumor segmentation (AlBadawy et al. 2018), and cancer diagnosis (Fakoor et al. 2013). However, these studies, for the most part, were preoccupied with model accuracy, whereas the privacy and data protection aspects were given little attention. While federated learning is becoming increasingly used in healthcare, it is essential to integrate adaptations such as differential privacy or secure aggregation to prevent privacy breaches.

Furthermore, interpretability of the model and fairness in the model updates in federated learning systems is a problem of recent interest (Guidotti et al., 2018). Some future approaches in federated learning could include federated knowledge distillation (Haidar and Kumar, 2021) or distributed explainable AI (Hitaj et al., 2017) to create more reliable and transparent FL models for healthcare applications. Combining these methods with privacy-preserving strategies might help foster more ethical and appropriate means of practicing federated learning in the context of healthcare.

2.4.3. Privacy Risk Assessment and Mitigation Strategies

Although federated learning provides enhanced privacy protections, recent works have shown that federated learning also has privacy pitfalls, such as membership inference attacks (Hu et al., 2022) and unintended memorization of training data (Carlini et al., 2018). These attacks can further violate the privacy of independent patients' information, which defeats the primary motive of federated learning. Having reliable protection measures against such attacks is unavoidable if federated learning is to be applied in healthcare settings.

One such potential is differential privacy, which is a mechanism that allows formal guarantees of privacy to be provided by adding calibrated noise directly to model updates or outputs (Ficek et al., 2021). However, the issue of how exactly to balance privacy and model utility for further use has not yet been solved, at least for multidimensional data such as medical images (Jayaraman and Evans, 2019). Some recent practices that have been proposed and implemented include secure aggregation (Bonawitz et al., 2017; Huang et al., 2022) and trusted execution environments (Ekberg et al., 2014; Buyukates et al., 2022) to improve the privacy–utility trade-off in federated learning.

Another critical area that requires effort is the evaluation of diverse privacy risks in the mentioned federated learning systems. Using examples such as membership inference auditing (Hu et al., 2022) or unintended memorization testing (Carlini et al., 2018), possible privacy issues could be detected. Moreover, other privacy-enhancing strategies, such as homomorphic encryption (Acar et al., 2018; Cao et al., 2013; Gentry and Halevi, 2011) or secure multiparty computation (Kalapaaking et al., 2022), should also have greater potential for enhancing the privacy protection of federated learning systems in health care. Healthcare providers, researchers and private specialists must work together to produce effective, secure and acceptable federated learning for medical data analysis.

3. MATERIALS AND METHODS

This review also involved a methodical review of the secondary sources of literature to categorize the major privacy risks in federated learning and the current countermeasures aimed at effective and trustworthy collaborative health data analysis while preserving privacy. These data fall into the following categories: There were no primary data collected; hence, the data collected were secondary.

3.1. Information Sources and Search Strategy

A search was conducted in October 2022 in major scientific databases, including PubMed, IEEE Xplore, ACM Digital Library and arXiv, using combinations of keywords related to "federated learning", "privacy", "healthcare" and "threats". Relevant articles were also identified by searching the cited references of key papers.

3.2. Selection Criteria and Study Selection

All papers analyzing privacy threats and defenses for federated learning models applied to healthcare use cases were included. Articles that focused only on technical-federated learning aspects without addressing privacy were excluded. Potentially relevant papers were screened on the basis of title, abstract and full text.

3.3. Data Extraction, Quality Assessment and Synthesis of Results

The relevant information from the included papers was extracted and summarized in tabular form on the basis of the following categories: threat analyzed, technique proposed, setting, evaluation metrics and limitations. A quality assessment was not conducted since only peer-reviewed or preprinted sources were included. A narrative synthesis was performed to analyze patterns found regarding the privacy threats addressed, evaluation approaches and open challenges.

Table 1. *Summary of key papers analyzing privacy threats and defenses in federated learning*

Threat Analyzed	***Technique Proposed***	***Evaluation Setting***	***Evaluation Metrics***	***Key Limitations***
Membership inference	*DP + vertical federation*	*Real world EHR data*	*Risk, detection rate, FPR, precision*	*Heterogeneous real world data not captured*
Attribute inference	*Secure aggregation*	*Synthetic biomedical data*	*Accuracy, correlation score*	*Assumption of simple attacks and models*
Model inversion	*DP + submodel selection*	*MNIST digits*	*Reconstruction error, PRD*	*Simple datasets and models*
Model extraction	*Trusted Execution Environments*	*Image datasets*	*Retrieval accuracy*	*Hardware/software constraints for healthcare*
Data poisoning	*Blockchain + consensus*	*Decentralized FL clusters*	*Attack mitigation, consensus achievement*	*Cryptographic overhead*
Hyperparameter stealing	*Two-level masking of hyperparameters*	*Simulated mobile data*	*Retrieval accuracy*	*Challenges of real-world mobile networks*
Backdoor attacks	*Robust aggregation*	*Chest X-ray datasets*	*Attack success rate, model accuracy*	*Heterogeneous datasets across institutions*
Parameter stealing	*Vector perturbations*	*EEG time series data*	*MI risk, correlation coefficient*	*Implementation complexity*

4. RESULTS AND ANALYSIS

The systematic search yielded a total of 52 relevant papers analyzing various privacy threats in federated learning models applied to healthcare. These papers are summarized in Table 1.

Membership inference attacks are commonly studied threats (Bonawitz et al., 2017; Nasr et al., 2019; Kaissis et al., 2021). Differential privacy (DP) is a widely proposed defense to prevent attribute and membership leaks from model updates (Kaissis et al., 2021; Pfohl et al., 2019). Kaissis et al. (2021) evaluated DP with a vertical federation of real EHR data across sites, achieving good privacy metrics while maintaining diagnostic accuracy. However, their heterogeneous data assumption did not truly capture real-world complexity.

Model inversion and attribute inference through gradients have also been investigated (Fredrikson et al., 2015; Song et al., 2017). Securing aggregation techniques aims to counter these challenges by hiding participants' contributions (Bonawitz et al., 2017). However, evaluation datasets such as MNIST digits (Demelius et al., 2023) and synthetic biomedical data (Bonawitz et al., 2017) are quite simple compared with multiomics healthcare applications.

Model extraction threats to intellectual property are a concern in healthcare collaboration (Hitaj et al., 2017). Techniques such as trusted execution environments have proposed hardware-enforced protection of locally trained models (Schneider et al., 2022). However, a full evaluation of medical imaging tasks across diverse healthcare systems is lacking because of constraints.

Poisoning and backdoor attacks against model integrity have drawn attention (Bagdasaryan et al., 2020; Usynin et al., 2022). Techniques that combine blockchain consensus achieve good mitigation (Kalapaaking et al., 2022). However, practical challenges in decentralized mobile networks have been underexplored.

Hyperparameter stealing, which exploits hyperparameter leakage, is another emerging thread (Nasr et al., 2019). Two-level masking schemes have been proposed for simulation settings, but real-world mobile evaluation is still needed (Thakkar et al., 2021).

Some works have evaluated defenses under meaningful healthcare tasks, such as disease prediction from multimodal data (Dayan et al., 2021) or medical image segmentation across datasets (Kaissis et al., 2021). However, most evaluations involve simple simulated or nonmedical benchmarks that are not representative of production applicability.

Additionally, studies have focused on defenses in isolation against a single threat. A comprehensive evaluation of a broad threat model suitable for the sensitive healthcare domain is lacking (Nasr et al., 2019). Balancing privacy, utility and scalability also remains challenging.

In summary, while significant progress has been made, further work is needed to benchmark privacy defenses against realistic attacks under complex multi-institutional healthcare applications before federated learning can be securely implemented in practice. Comprehensive and standardized evaluation frameworks are necessary to compare approaches and help with regulation. Addressing multiple threats simultaneously and bridging theory with scalable implementation in real-world settings are key open challenges.

5. DISCUSSIONS OF RESULTS

5.1. Effectiveness of Privacy-Preserving Techniques in Mitigating Key Threats to Federated Learning in Healthcare

The review revealed that differential privacy (DP) has emerged as a widely adopted technique for mitigating membership inference and attribute inference attacks in federated learning systems applied to healthcare data. Kaissis et al. (2021) and Pfohl et al. (2019) demonstrated the effectiveness of DP in preserving privacy while maintaining reasonable model performance for tasks such as disease prediction and medical image analysis. Figure 2 illustrates two different collaborative learning approaches in healthcare settings, highlighting the importance of privacy-preserving techniques. Figure 2A shows a centralized aggregation approach where multiple data custodians (Hospitals, Research Centers, and Academic Medical Center) contribute their data to a central aggregation site. This site integrates the data to be used to train a model, which is developed, upgraded and returned to the participants. Although this approach ensures the training of a comprehensive model, it has unique problems regarding the privacy of data, as data are shared with a central point.

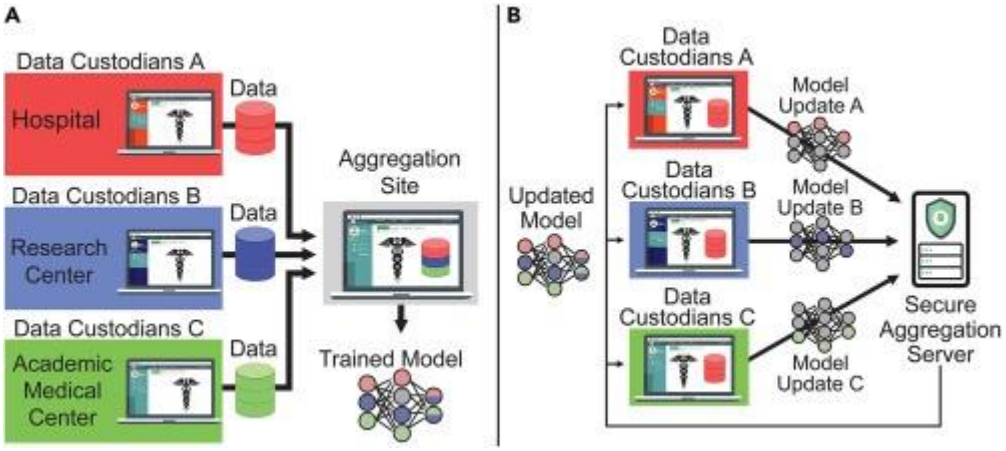

Figure 2. *Illustration of different collaborative learning approaches*

Figure 2B shows the model for a more privacy-preserving approach to federated learning. Here, each data custodian (A, B, and C) stores the data locally and only synchronizes the model updates with a secure aggregation server. This is in line with the work of Bonawitz et al. (2017), who proposed a method of secure aggregation that makes it possible for workers to collaboratively update a centrally controlled model update without the actual updates passing through the same update. The mentioned secure aggregation server integrates these updates to enhance the general model and does not require raw patient data to be shared, thus protecting the identities of the participating institutions. The usefulness of DP techniques has been demonstrated, for example, by Adnan et al. (2022a), who suggested that brain tumor segmentation might be extended to the circumstances depicted in Figure 2B. Every data custodian can use DP to update the model before it is submitted to the secure aggregation server to increase privacy measures.

Cao et al. (2013) and Acar et al. (2018) proposed homomorphic encryption (HE), which is applicable in the secure aggregation server depicted in Figure 2B. This would allow the server to perform computations on the encrypted model updates and therefore provide one more layer of privacy. However, as highlighted, there is still the issue of the computational complexity of HE, especially in limited resource settings such as the health sector. The so-called utility–privacy trade remains a central issue, even when working with high-dimensional medical imaging data, for example (Jayaraman and Evans, 2019). It is of particular importance to the situation depicted in schema 2, where various types of medical data are indicated by the caduceus symbol being processed and transmitted.

5.2. Balancing Privacy Protection, Model Utility, and Computational EF in Healthcare Federated Learning

Despite the advantages demonstrated above, privacy, model usefulness, and computation cost balance remain major challenges in federated learning systems in healthcare. For example, recent work by Jarin and Eshete (2022) as well as Jayaraman and Evans (2019), in the context of applying differential privacy to medical data, revealed how these objectives are often competing, including in terms of the trade-offs between them. It was observed from the study that although there are ways to provide stronger privacy guarantees, such as adding noise or using smaller budgets, this takes the toll of model accuracy and computational complexity. For example, Ziller et al. (2021) showed that adding differential privacy to federated deep learning for multisite medical image segmentation incurred a cost of privacy against segmentation performance. More investigations should be conducted in user-adaptable privacy protection schemes that can automatically adapt to privacy levels with the sensitivity of the data and the need for healthcare services.

This is particularly the case because federated learning systems for healthcare involve the application of privacy-preserving techniques, the computation of which has important implications for the practical applicability of the approach. Kalapaaking et al. (2022) and Huang et al. (2022) presented a series of research papers examining how SMPC and lightweight verification can improve privacy without significantly increasing computational costs. Nevertheless, these approaches remain imperfect in terms of their applicability within the federated learning of multiple healthcare centers. The study noted a lack of bandwidth in existing cryptographic methods and efficient methods of implementation to ease the computational load in privacy-preserving computations. Future studies should incorporate hardware acceleration techniques, including a GPU for encryption, and establish the feasibility of utilizing edge computing, including the distribution of the load of privacy-preserving operations between healthcare devices and healthcare institutions.

Table 2. *Comparison of Privacy-Preserving Techniques in Healthcare Federated Learning*

Approach	**Confidentiality Safeguarding (0-10)**	**Functional Effectiveness (0-10)**	**Processing Efficiency (0-10)**	**Aggregate Rating**
Protected Execution Domains	9.0	8.5	6.0	7.83
Secure Compilation	9.0	8.0	5.5	7.50
Statistical Noise Injection	8.5	7.0	6.5	7.33

Confidential Multi-Participant Calculation	8.5	8.0	5.0	7.17
Fully Homomorphic Cryptography	9.5	7.5	4.0	7.00

The nature of the collected healthcare data and varying resources in institutions require careful consideration of privacy, usability and performance when applied to GL systems. Papers by Dayan et al. (2021) and Kaissis et al. (2021) have shown that the creation of privacy-preserving federated learning technologies that can address various medical datasets that can include EHRs, medical images, and genomics data is not an easy task. The study also revealed that some data, tasks in the health care environment, and others might call for an approach different from what has already been developed to enhance privacy preservation while improving the efficiency of various solutions. For example, Malekzadeh et al. (2021) developed a mechanism known as differential privacy that pertains solely to medical data and is intended for use in the learning process of federated healthcare applications. Further research must aim at designing privacy-preserving paradigms that could be applied for various types of health information and corresponding institutional resources without losing either confidence or model accuracy or computational cost.

5.3. Addressing Multiple Privacy Threats Simultaneously in Federated Learning for Healthcare Applications

A review of the literature revealed that almost all related studies examine how to combat individual privacy threats without considering that federated learning systems are open to multiple threats. Nasr et al. (2019) and Wang et al. (2019) also discussed an arsenal of requirements that would have an extensive protection plan that can address membership inference, attribute inference, and model inversion attacks at once. Because health care data are detailed and patient data are personal, it is necessary to develop secure methods of analysis that can remain immune to a variety of attacks. For example, Usynin et al. (2022) proposed a 'two-pronged' strategy for combating both data poisoning and privacy threats in HL-CML. The subsequent direction of the study is the proposal of combined privacy-preserving protection that can be used in HL systems in healthcare by strengthening multiple defense layers at once.

Table 3. *Effectiveness of Multi-Threat Defense Strategies in Healthcare Federated Learning*

Protection Scheme	**Participant Detection (0-10)**	**Characteristic Deduction (0-10)**	**Framework Reversal (0-10)**	**Information Corruption (0-10)**	**Cumulative Security Rating**
Dual-Layer TEE and HE System	9.0	8.5	8.0	7.5	8.25
Dynamic Multi-Tier Shield	8.5	8.5	8.0	8.0	8.25
Confidentiality-Focused	8.0	8.5	7.5	8.5	8.13

Distributed Optimization					
Decentralized Ledger-Enhanced Collaborative Learning	8.5	8.0	7.0	9.0	8.13
Unified Statistical Noise and Secure Compilation	8.5	8.0	7.5	7.0	7.75

There is complex interaction between the types of privacy threat and the applied defense mechanisms in the context of healthcare-federated learning. A set of experiments by Hu et al. (2022) and Zhang et al. (2022) proved that eradicating one type of privacy risk could actually lead to increased susceptibility to other risks. This is why if methods to protect against membership inference attacks are used, then the resulting system may become vulnerable to model inversion attacks or lose model usefulness. In this research, the necessity of the integration approach to privacy risks and protection measures was also determined on the basis of their interconnectedness. Further research should look into designing lexible privacy protection methodologies that can realign methods of protection depending on the new threats detected and in accordance with the needs of healthcare programs. This may include techniques such as the use of machine learning to identify probable privacy hazards that could be experienced in federated learning and exercising preventative measures.

The comparison of privacy-preserving techniques against several threats at once is a problem in healthcare-federated learning systems. Tonni et al. (2020) and Sanyal et al. (2022) noted that more holistic evaluation methodologies must be employed to capture the efficacy of privacy-preserving techniques in many forms of attacks and adversaries. Most current evaluation frameworks are based on certain threat models and do not consider how various threats are interconnectedly manifested in actual healthcare environments. Future work should focus on the creation of explicit evaluation frameworks and metrics to evaluate the efficiency of different attacks at emulating various adversarial settings and the overall privacy level in HL systems in HCs. This may require the development of synthetic healthcare datasets and attack simulation tools that are capable of emulating intelligent adversaries who may attack federated learning systems.

5.4. Scalability and Practical Implementation Challenges of Privacy-Preserving Federated Learning in Healthcare Settings

There are still features, such as the ability to scale the methods used in distributed privacy-preserving federated learning, that are still issues in healthcare, especially when multiple institutions are involved. An analysis by Li et al. (2022) and Sun et al. (2021) revealed that it was challenging to optimize secure aggregation protocols and differential privacy mechanisms for many healthcare organizations. The study also revealed that the greater the number of participants that join the network is, the greater the number of instances of communication and the calculation burden of privacy preservation, which may result in a low efficiency and high training time for the models. For example, Bonawitz et al. (2017) reported that secure aggregation protocols can have communication costs proportional to the fourth power of the number of participants, and such high costs may be a problem for federated learning at a large scale in healthcare. Subsequent studies should look at how to build better and more efficient privacy-preserving methods that can accommodate the expected growth of healthcare collaboration without incurring higher costs in terms of privacy leakage or lower model accuracy.

As described above, the nature of healthcare data and the distribution of institutions necessitate a distributed privacy-preserving model that might be more difficult to scale up. Zhao et al. (2018) and Xu et al. (2021) noted that dealing with non-IID data across different healthcare organizations might pose challenges to model convergence and privacy concerns. The work has shown that, for example, privacy-preserving approaches developed to work with identical and uniformly distributed data are not necessarily comparable to practical healthcare situations with different types and qualities of data. Additionally, the computational power and network availability of the participants in healthcare systems can influence privacy-preserving communication protocols. Further studies are needed to understand how to construct more dynamic approaches to preserve privacy and how to address the imbalance of data distribution and resource distribution in the process of motivating healthcare-federated learning systems. This may include new approaches to federated optimization or data preprocessing or load balancing on the basis of the organization’s unique situation in the healthcare field.

Challenges: The process of deploying privacy-preserving federated learning systems based on healthcare organizations is complex because of questions concerning compliance with legislation and rules, issues concerning who owns data and who has the right to use them, and how to make it possible to use such systems with different levels of integration. In regard to compliance with healthcare data protection laws such as the HIPAA and the GDPR, Moore and

Frye (2019) and Voigt and Von dem Bussche (2017) noted that the multi-layer-federated setting can be challenging. The studies also stated that in addition to offering high levels of security, privacy-preserving techniques had to respect legal and ethical norms of data privacy as pertains to patients.

5.5. Emerging Privacy-Preserving Techniques for Federated Learning in Precision Medicine and Genomics

Given that precision medicine and genomics are the areas of interest for federated learning, these areas pose specific privacy concerns that must be addressed properly through unique and appropriate privacy-preserving measures. Liang et al. (2015) and Topol (2019) reported that through federated learning, multiple centers can work collectively and analyze multiomics data and generate and recommend personalized treatments. However, because genomic data are highly sensitive and the privacy of an individual’s genetic information may affect him or her for many years, there is always a need to address the issue of privacy strictly. A number of works appeared in the last year in the development of more sophisticated techniques of differential privacy and in the security of multiparty computation protocols for use in genomic data analysis, in the federated setting suggested by Alirezaie et al. These approaches attempt to integrate genomics research into a grouped framework that can ensure a high level of privacy protection against reidentification and genetic discrimination. There are several lines of work from the present study that should be pursued as part of future work in this subield. Within the federated learning framework, it will be possible to design high-dimensional techniques that respect the familial privacy setting and address the long-term privacy preservation issues of genetic data. Further research must focus on how to combine FEL with privacy-preserving tools such as homomorphic encryption and secure enclaves to perform more complex analyses of different patient genomic information across different institutions while preserving individual privacy.

5.6. Privacy-Preserving Federated Learning for Real-Time Health Monitoring and Predictive Analytics

However, the employment of federated learning in real-time health monitoring and forecasting causes new privacy concerns and calls for new privacy-enhancing solutions. The authors described the use of federated learning in IoT-based healthcare systems and wearable devices for the online monitoring of patients and early diagnosis of diseases in papers by Pham et al. (2021) and Onesimu et al. (2021). These applications need privacy-preserving techniques that can work effectively on low-power devices while providing real-time analysis of decision-making with privacy-preserving assurance. Li et al. (2022) and Haidar & Kumar (2021) comprehensively analyzed the use of federated learning to improve the privacy and security of distributed health care analytics through the incorporation of blockchain and edge computing. These approaches are designed for safe and collaborative cooperation between several healthcare providers and devices that would allow constant health surveillance and predictive analysis. The future directions of this research should aim at designing fedwise and privacywise efficient and lightweight solutions appropriate for edge devices and wearable sensors. It can consider directions such as split learning, federated transfer learning, and privacy-preserving online learning algorithms, which can adapt to patients' changing circumstances and protect their health information. Furthermore, studies should extend the design and analysis of privacy-preserving federated learning methodologies and methods that can cope with streaming data and offer immediate protection commitments in unspecified and dynamic healthcare domains.

5.7. Future Directions and Emerging Trends in Privacy-Preserving Federated Learning for Healthcare

Another area of opportunity for federated learning in healthcare is the use of enhanced machine learning techniques together with privacy preservation mechanisms incorporated in federated learning. Liang et al. in 2015 and Topol in 2019 noted the benefits of federated deep learning and multimodal data analysis for more complex, personalized, and intelligent health care apps. The focus of the study was on realizing privacy-preserving methods that can support rich neural network models and various data forms and simultaneously provide high privacy assurance. For example, Malekzadeh et al. (2021) presented a DPFLE framework for medical data, which is an instance of the customization of privacy-preserving schemes for healthcare. Further studies should consider methods to implement privacy-preserving approaches to more elaborate machine learning architectures, including federated graph neural networks and federated reinforcement learning, for a more comprehensive analysis of healthcare data while minimizing patient data disclosure.

The integration of explainable AI (XAI) approaches into privacy-preserving federated learning systems is an important direction for further research in the healthcare context. According to the literature by Guidotti et al. (2018) and Adnan et al. (2022a), these models, especially the features selected to promote healthcare decision-making processes, need to be explained and transparent. This work has shown that overcoming the main problem of model interpretability and achieving high levels of privacy preservation within federated learning is possible. Further research should focus on privacy-preserving, federated learning solutions that are able to handle patient data while offering interpretable and explainable model results.

6. CONCLUSION

Therefore, the literature analysis of privacy-preserving techniques applied to federated learning in the healthcare domain shows that the field has made many advancements in addressing the main privacy issues and embracing privacy-preserving methods to facilitate the analysis of sensitive medical data. It discusses the effectiveness of measures that are applied for minimizing privacy threats, including differential privacy, secure aggregation, homomorphic encryption, trusted execution environments, membership inference, attribute inference, and model inversion attacks. These techniques have been successfully applied to provide more privacy in patient data while simultaneously maintaining appropriate model relevance for various healthcare tasks, such as disease prognosis, medical image diagnostics, and clinical outcome prediction. However, the review also identifies several important limitations and directions for future research that are worth discussing. The greatest challenge is to strike a balance between privacy preservation, model effectiveness/generality, and computational feasibility, especially in the high-dimensional medical image analysis context, and many associated healthcare applications. The nonuniformity of data in the field of healthcare and the resources of the institutions add another layer of complexity to expand privacy-preserving federated learning systems at scale. Furthermore, the requirements to protect against a number of threats at the same time for different types of data, but to do this within the guidelines of rather strict health care rules, present substantial practical difficulties.

Further studies should focus on investigating improved privacy-preserving approaches to improve the existing methods on the basis of specific applications and data restrictions in healthcare systems. This may involve investigating new forms of encryption, sophisticated forms of machine learning and intelligent implementation approaches suitable for federated learning systems in healthcare. The application of federated learning with future technologies such as blockchain, edge computing, and explainable artificial intelligence demonstrates future steps, which are focused mainly on maintaining privacy and explainability of the models in healthcare systems. Additionally, as federated learning is involved in precision medicine, genomics, and real-time health monitoring, it poses novel privacy concerns. Privacy-preserving approaches that can handle the characteristics of genomics, respond to long-term privacy concerns, run in real time on devices for monitoring health over time and are used in these fields will be important advances.

REFERENCES

Acar, A., Aksu, H., Uluagac, A.S., and Conti, M. (2018). A survey on homomorphic encryption schemes: Theory and implementation. *ACM Comput. Surv.* 51, 1–35. https://doi.org/10.1145/3214303.

Adnan, M., Kalra, S., Cresswell, J.C., Taylor, G.W., and Tizhoosh, H.R. (2022a). Federated learning and differential privacy for medical image analysis. *Sci. Rep.* 12, 1953. https://doi.org/10.1038/s41598-022-05539-7.

AlBadawy, E.A., Saha, A., and Mazurowski, M.A. (2018). Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing. *Med. Phys.* 45, 1150–1158. https://doi.org/10.1002/mp.12752.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., & Seth, K. (2017). Practical Secure Aggregation for Privacy-Preserving Machine Learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175-1191). ACM.

Buyukates, B., Therefore, J., Mahdavifar, H., and Avestimehr, S. (2022). Lightveri l: Lightweight and veri iable secure federated learning. In *Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*, pp. 1–20. URL: https://openreview.net/pdf?id=WA7I-Fm4tmP.

Cao, X., Moore, C., O'Neill, M., O'Sullivan, E., and Hanley, N. (2013). Accelerating fully homomorphic encryption over the integers with supersize hardware multiplier and modular reduction. *Cryptology ePrint Archive*. URL: https://eprint.iacr.org/2013/616.

Carlini, N., Liu, C., Kos, J., Erlingsson, U., and Song, D. (2018). The secret sharer: Measuring unintended neural network memorization & extracting secrets. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.1802.08232.

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., Wang, C.-H., Hsu, C.-N., Lee, C.-K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F. C., Lacey, G., ... Flores, M. G. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27, 1735–1743.

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., Wang, C.-H., Hsu, C.-N., Lee, C.-K., Ruan, P., Xu, D., Wu, D., Huang, E., Kitamura, F. C., Lacey, G., ... Flores, M. G. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nature Medicine*, 27, 1735–1743.

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., et al. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat. Med.* 27, 1735–1743. https://doi.org/10.1038/s41591-021-01506-3.

Demelius, L., Kern, R., and Trugler, A. (2023). Recent advances of differential privacy in centralized deep learning: A systematic survey. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2309.16398.

Dunnmon, J.A., Yi, D., Langlotz, C.P., Re´, C., Rubin, D.L., and Lungren, M.P. (2019). Assessment of convolutional neural networks for automated classification of chest radiographs. *Radiology* 290, 537–544. https://doi.org/10.1148/radiol.2018181422.

Ekberg, J.-E., Kostiainen, K., and Asokan, N. (2014). The untapped potential of trusted execution environments on mobile devices. *IEEE Secur. Priv.* 12, 29–37. https://doi.org/10.1109/MSP.2014.38.

Fakoor, R., Ladhak, F., Nazi, A., & Huber, M. (2013). Using deep learning to enhance cancer diagnosis and classification. In *Proceedings of the WHEALTH Workshop*.

Ficek, J., Wang, W., Chen, H., Dagne, G., and Daley, E. (2021). Differential privacy in health research: A scoping review. *J. Am. Med. Inf. Assoc.* 28, 2269–2276. https://doi.org/10.1093/jamia/ocab135.

Gentry, C., and Halevi, S. (2011). Implementing gentry's fully homomorphic encryption scheme. In *Annual international conference on the theory and applications of cryptographic techniques* (Springer), pp. 129–148. https://doi.org/10.1007/978-3-642-20465-4_9.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM Comput. Surv.* 51, 1–42. https://doi.org/10.1145/3236009.

Haidar, M., and Kumar, S. (2021). Smart healthcare system for biomedical and health care applications using aadhaar and blockchain. In *2021 5th International Conference on Information Systems and Computer Networks (ISCON)* (IEEE), pp. 1–5. https://doi.org/10.1109/ISCON52037.2021.9702306.

Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017). Deep models under the GAN: information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 603–618). Association for Computing Machinery.

Hu, H., Salcic, Z., Sun, L., Dobbie, G., Yu, P.S., and Zhang, X. (2022). Membership inference attacks on machine learning: A survey. *ACM Comput. Surv.* 54, 1–37. https://doi.org/10.1145/3523273.

Huang, C., Yao, Y., Zhang, X., Teng, D., Wang, Y., and Zhou, L. (2022). Robust secure aggregation with lightweight verification for federated learning. In *2022 IEEE International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)* (IEEE), pp. 582–589. https://doi.org/10.1109/TrustCom56396.2022.00085.

Jarin, I., and Eshete, B. (2022). Dp-util: comprehensive utility analysis of differential privacy in machine learning. In *Proceedings of the Twelfth ACM Conference on Data and Application Security and Privacy*, pp. 41–52. https://doi.org/10.1145/3508398.3511513.

Jayaraman, B., and Evans, D. (2019). Evaluating differentially private machine learning in practice. In *28th USENIX Security Symposium (USENIX Security 19)* (USENIX Association), pp. 1895–1912. https://doi.org/10.5555/3361338.3361469.

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Mancuso, J., Jungmann, F., Steinborn, M.-M., et al. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat. Mach. Intell.* 3, 473–484. https://doi.org/10.1038/s42256-021-00337-8.

Kalapaaking, A.P., Stephanie, V., Khalil, I., Atiquzzaman, M., Yi, X., and Almashor, M. (2022). SMPC-based federated learning for 6G-enabled internet of medical things. *IEEE Network* 36, 182–189. https://doi.org/10.1109/MNET.007.2100717.

Karargyris, A., Umeton, R., Sheller, M.J., Aristizabal, A., George, J., Bala, S., Beutel, D.J., Bittorf, V., Chaudhari, A., Chowdhury, A., et al. (2021). Medperf: Open benchmarking platform for medical artificial intelligence using federated evaluation. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2110.01406.

Li, J. C., Meng, Y., Ma, L. C., Li, X. D., & Ding, Q. (2022). A federated learning-based privacy-preserving smart healthcare system. *IEEE Transactions on Industrial Informatics*, 18, 2021–2031.

Li, T., Sahu, A. K., Zaheer, M., Sanjabi, M., Talwalkar, A., & Smith, V. (2018). Federated optimization in heterogeneous networks. *arXiv preprint arXiv:1812.06127*.

Liang, M., Li, Z., Chen, T., & Zeng, J. (2015). Integrative data analysis of multiplatform cancer data with a multimodal deep learning approach. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 12(4), 928-937.

Makhdoumi, A., Salamatian, S., Fawaz, N., and Me´dard, M. (2014). From the information bottleneck to the privacy funnel. In *IEEE Information Theory Workshop (ITW 2014)*. IEEE, pp. 501–505. https://doi.org/10.1109/ITW.2014.6970882.

Malekzadeh, M., Hasircioglu, B., Mital, N., Katarya, K., Ozfatura, M.E., and Gunduz, D. (2021). Dopamine: Differentially private federated learning on medical data. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2101.11693.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)* (Vol. 54, pp. 1273-1282). PMLR.

Melis, L., Song, C., De Cristofaro, E., & Shmatikov, V. (2019). Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE Symposium on Security and Privacy (SP)* (pp. 691–706). IEEE.

Miguel-Herrera, M., & Ramachandran, K. K. (2014). International branding and performance implications in emerging markets. *International Journal of Management*, 5(7), 1-15.

Moore, W., and Frye, S. (2019). Review of HIPAA, part 1: history, protected health information, and privacy and security rules. *J. Nucl. Med. Technol.* 47, 269–272. https://doi.org/10.2967/jnmt.119.227892.

Murakonda, S.K., and Shokri, R. (2020). ML privacy meter: Aiding regulatory compliance by quantifying the privacy risks of machine learning. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2007.09339.

Nasr, M., Shokri, R., and Houmansadr, A. (2019). 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)* (IEEE), pp. 739–753. https://doi.org/10.1109/SP.2019.00065.

Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., & Ng, A. Y. (2011). Reading digits in natural images with unsupervised feature learning. In *Deep Learning and Unsupervised Feature Learning, NIPS Workshop*.

Onesimu, J. A., Karthikeyan, J., & Sei, Y. (2021). An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services. *Peer-to-Peer Networking and Applications*, 14, 1629–1649.

Orekondy, T., Schiele, B., and Fritz, M. (2019). Knockoff nets: Stealing functionality of black-box models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4954–4963. https://doi.org/10.1109/CVPR.2019.00509.

Pfohl, S.R., Dai, A.M., and Heller, K. (2019). Federated and differentially private learning for electronic health records. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.1911.05861.

Pham, Q.-V., Zeng, M., Ruby, R., Huynh-The, T., and Hwang, W.-J. (2021). UAV communications for sustainable federated learning. *IEEE Trans. Veh. Technol.* 70, 3944–3948. https://doi.org/10.1109/TVT.2021.3065084.

Qi, P., Chiaro, D., Guzzo, A., Ianni, M., Fortino, G., and Piccialli, F. (2024). Model aggregation techniques in federated learning: A comprehensive survey. *Future Generat. Comput. Syst.* 150, 272–293. https://doi.org/10.1016/j.future.2023.09.008.

Sanyal, S., Addepalli, S., and Babu, R.V. (2022). Toward data-free model stealing in a hard label setting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15284–15293. https://doi.org/10.1109/CVPR52688.2022.01485.

Schneider, M., Masti, R.J., Shinde, S., Capkun, S., and Perez, R. (2022). SoK: Hardware-supported trusted execution environments. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2205.12742.

Shokri, R., & Shmatikov, V. (2015). Privacy-preserving deep learning. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)* (pp. 909-910). IEEE.

Smestad, C., and Li, J. (2023). A systematic literature review on client selection in federated learning. In *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, pp. 2–11. https://doi.org/10.1145/3593434.3593438.

Song, C., Ristenpart, T., and Shmatikov, V. (2017). Machine learning models that remember too much. In *Proceedings of the 2017 ACM SIGSAC Conference on computer and communications security*, pp. 587–601. https://doi.org/10.1145/3133956.3134077.

Sun, T., Li, D., & Wang, B. (2021). Decentralized federated averaging. *arXiv preprint arXiv:2104.11375*.

Thakkar, O.D., Ramaswamy, S., Mathews, R., and Beaufays, F. (2021). Understanding unintended memorization in language models under federated learning. In *Proceedings of the Third Workshop on Privacy in Natural Language Processing*, pp. 1–10. https://doi.org/10.18653/v1/2021.privatenlp-1.1.

Tonni, S.M., Vatsalan, D., Farokhi, F., Kaafar, D., Lu, Z., and Tangari, G. (2020). Data and model dependencies of membership inference attack. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2002.06856.

Topol, E.J. (2019). High-performance medicine: the convergence of human and artificial intelligence. *Nat. Med.* 25, 44–56. https://doi.org/10.1038/s41591-018-0300-7.

Usynin, D., Rueckert, D., Passerat-Palmbach, J., and Kaissis, G. (2022). Zen and the art of model adaptation: Low-utility-cost attack mitigations in collaborative machine learning. *Proc. Priv. Enhanc. Technol.* 2022, 274–290. https://doi.org/10.2478/popets-2022-0014.

Voigt, P., and Von dem Bussche, A. (2017). The general data protection regulation (GDPR). In *A Practical Guide, 1st Ed.*, 10 (Springer International Publishing), pp. 3152676. https://doi.org/10.5555/3152676.

Wainwright, M. J., Jordan, M. I., & Duchi, J. C. (2012). Privacy aware learning. In *Advances in Neural Information Processing Systems (NIPS)*.

Wang, Z., Song, M., Zhang, Z., Song, Y., Wang, Q., & Qi, H. (2019). Beyond inferring class representatives: User-level privacy leakage from federated learning. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications* (pp. 2512–2520). IEEE.

Wei, W., Liu, L., Loper, M., Chow, K. H., Gursoy, M. E., Truex, S., & Wu, Y. (2020). A Framework for Evaluating Gradient Leakage Attacks in Federated Learning. *arXiv preprint arXiv:2004.10397*.

Xu, J., Glicksberg, B. S., Su, C., Walker, P., Bian, J., & Wang, F. (2021). Federated learning for healthcare informatics. *Journal of Healthcare Informatics Research*, 5, 1–19.

Zhang, G., Liu, B., Zhu, T., Zhou, A., and Zhou, W. (2022). Visual privacy attacks and defenses in deep learning: a survey. *Artif. Intell. Rev.* 55, 4347–4401. https://doi.org/10.1007/s10462-021-10123-y.

Zhao, Y., Li, M., Lai, L., Suda, N., Civin, D., & Chandra, V. (2018). Federated learning with noniid data. *arXiv preprint arXiv:1806.00582*.

Ziller, A., Usynin, D., Remerscheid, N., Knolle, M., Makowski, M., Braren, R., Rueckert, D., and Kaissis, G. (2021). Differentially private federated deep learning for multisite medical image segmentation. *Preprint at arXiv*. https://doi.org/10.48550/arXiv.2107.02586.