# Harnessing Federated Learning for Secure Data Sharing in Healthcare Systems

Salih Mahmoud Attya
Alnoor University
Nineveh, Iraq
salih.mahmoud@alnoor.edu.iq

Doaa Ali Jumaa
Al Mansour University College
Baghdad, Iraq
doaa.ali@muc.edu.iq

Noora Nazar Kamal Hwezy
Al Hikma University College
Baghdad, Iraq
Nora.nazar@hiuc.edu.iq

Wafaa Adnan Sajid
Al-Rafidain University College
Baghdad, Iraq
wafa@ruc.edu.iq

Ali Mohammed Khaleefah
Al-Turath University
Baghdad, Iraq
alimohamed@turath.edu.iq

Genadiy Zhyrov
Taras Shevchenko National University of Kyiv
Kyiv, Ukraine
genadiy.zhyrov@knu.ua

Haider Ali
Uruk University
Baghdad, Iraq
Haider_ali@uruk.edu.iq

*Abstract*— **Background: The digitization of healthcare data has made significant progress in medical research and personalized medicine. Nonetheless, conventional centralized data-sharing structures present primary obstacles to information privacy and security due to laws like the Health Insurance Portability and Accountability Act (HIPAA). Federated Learning (FL) has been proposed as a potential solution that can enable collaborative learning with decentralized datasets without the requirement of data centralization.**

**Objective: This study examines how well Federated Learning performs in meeting both the requirements of secure data exchange between healthcare organizations and achieving high model accuracy without violating any privacy compliance regulations.**

**Methods: A Federated Learning framework was implemented with a neural network model using federated learning on an electronic health records (EHR) database collected from multiple hospitals. Its accuracy was compared to a traditional centralized model across various states, while the people also consider its convergence speed and data leakage risks. We incorporated differential privacy mechanisms in order to improve the security of data sets as well as prevent malicious attacks.**

**Results: The FL model achieved accuracy comparable to the centralized model, with only a marginal reduction. Furthermore, the integration of differential privacy significantly reduced the risk of data breaches, providing robust protection against adversarial attacks.**

**Conclusion: The FL model reported a performance difference which was only slightly reduced as compared with the centralized model. Moreover, this privacy compliance overcomes the risk of data breaches by integrating with differential privacy & securing against adversarial attacks.**

## I. INTRODUCTION

A major transformation in the healthcare industry has been witnessed due to the widespread adoption of electronic health records (EHRs) which makes it easier and more accurate to access, control and analyze patient data. A big source of useful data is EHR, which contains several patient-related information — customized health care to prediction analysis.

However, the storing and centralizing of personal data is put under strain because of privacy, security concerns as well as legislation, such as the Health Insurance Portability and Accountability Act (HIPAA). This exacerbates these fears, particularly when those models of data sharing involve the roll-up of patient records from lots of different institutions into one central database. Today, the risks of data breaches, unauthorized access, and misuses of patient information are not lower than at any moment before, actually healthcare is seeing an increase in incidents due to the high value associated with this type of target facing cybercriminal threats. Recent studies show cyberattacks on healthcare data are increasing in complexity, reinforcing the importance of decentralized and privacy-preserving ways to share that data, such as Federated Learning [FL). FL addresses these privacy concerns and also maintains strong model performance by allowing collaborative learning amongst institutions without requiring raw data to move back-and-forth [1], [2].

Importantly, this has a serious risk of data breaches, signifying the need to share health information quickly with increased security [3], [4]. Indeed, the number and sophistication of cyber-attacks affecting health systems have only worsened [5]. One can see in this perspective a growing need for secure data exchange platforms, to guarantee privacy and at the same time facilitate collaborative research with patients' personal information being intentionally or unintentionally exposed. FL has been considered as one of the most attractive decentralized solutions to these problems [2], [6].

Federated Learning allows organizations to collaboratively train models on a large corpus of decentralized data—all without exchanging the raw dataset. Since the retrieval and training of FL frameworks can be applied to healthcare scenarios for data privacy while testing, an advantageous model. One example of this is the secure management of electronic health records (EHR) while keeping sensitive data privacy was shown by Salim and Park in [3] using advanced

encryption techniques in FL. Patel et al. [6] examine applications of FL and its use for multi-institutional healthcare collaboration

Instead, each site trains a model using its data and shares only the updates to their gradients with a central server. A model: By the server, joint updates of parameters to an aggregated model [4], [7]. Federated learning (FL) has achieved success in several sectors like financial, telecom, and recently healthcare [1], [5].

Among them, FL has been proposed as an effective algorithmic tool for privacy in health care. Xu et al. sheds some light on how FL use can be beneficial for healthcare informatics where data does not need aggregation at one place and doesn't require centralized approach to collect all the data. Further, FL has been shown to be very effective in reducing risks of data leakage using advanced privacy enhancing techniques such as differential privacy and secure multi-party computation [2], [5].

This is important in healthcare due to adherence of privacy regulations and ability for creating customizable models without reliance on ownership of the data [3], [6]. The healthcare sector might face certain challenges to implement FL which goes beyond any understanding and associated benefits the same may bring in. Variations in the data types across different sets may affect model training and performance, and pose difficulties to an implementer. Data of value and percentage differences affect model performance it can lead to biases while generalizing the behavior [8], [9]. In addition, the collaboration decreases with FL reducing the need to directly transfer data from server to client in raw form, however preserving privacy is still a challenge.

While the presence of adversarial techniques has led to advancements in this area, one major concern that still persists with model-sharing is privacy risk [10], [11]. Therefore, researchers are striving to combine more privacy-enhancing technologies like differential privacy and secure multi-party computation into the FL framework for better data protection [12], [11].

Today we see an article on Federated Learning tackling the hard problem of data sharing in healthcare by being secure and efficient. Specifically, this work is to understand can FL model the data and still comply with privacy regulations. We also investigated to what degree employing differential privacy techniques has an impact on model performance as well as data security, and evaluated whether Federated Learning could provide a way forward instead of traditional centralized options for sharing data in the healthcare industry.

The study adds to the increasingly vast literature on FL and its applications by addressing fundamental concepts around privacy, and security-related issues; providing a crucial perspective into how is it done in practice along with implications arising from this [6], [7], [9].

### *A. The Study Objectives*

The article will explore how Federated Learning could serve as a secure and privacy- safe approach to distributing data within healthcare systems. As electronic health record (EHR) use continues to rise and collaborative research across institutions becomes increasingly in demand, traditional data-sharing methods have substantial difficulty striking a balance between facilitating adequate speed of access and maintaining strict patient privacy protections. This study hopes to overcome these challenges and aims to examine the promise of federated learning — enabling more sophisticated machine learning models while preventing the centralization of private patient data. In particular, the article aims to examine in what manner FL can ensure both accurate models and their availability access with lower risks of data breaches and unauthorized actions. Also, performed an analysis of integrating differential privacy methods into the FL system to see how this improves on security and confidentiality of decentralized data. By focusing particularly on healthcare, this paper demonstrates that FL has the potential to transform data-sharing in the health sector as well and — with the deployment of a novel framework incorporated within the existing infrastructural setup which its performance is analyzed here for can foster more privacy-preserving and wide-spread research-driven services across different clinics/hospitals.

### *B. Problem Statement*

The ever-increasing digitization of evidence in healthcare has resulted in more and more necessitating electronic health records (EHRs) within medical organizations. This digital transformation has the potential to revolutionize healthcare, from improving patient care and health outcomes to accelerating important medical research. So, these things are great to work with but at the same time, they come along with a really big hassle of data privacy where there are some regulations like GDPR and HIPAA! Even when consolidating patient data, as one might do with how healthcare has traditionally exchanged information as well, you run significant risks like increased chances of unauthorized access to personal medical records and possible abuse. Furthermore, due to the heterogeneity of data types across different institutes, it is problematic to find more generalizable and novelty-preserving models that incorporate multi-source granular-level (i.e., local) relational association rules as the global constraints on privacy and legally-sensitive information disclose how they were derived.

A major bottleneck in this pandemic outbreak that has been rooted are struggling with is how multiple health organizations could collaborate to share patient data-powered research without compromising the privacy and security of information. The rise of cybersecurity threats and data governance requirements has rendered the traditional centralized models for information exchange neither secure nor scalable. Furthermore, there is an important need for data analytical techniques that are capable of handling the numerous yet disorganized aspects of healthcare datasets and at the same time able to not jeopardize precision or effectiveness in creating new models.

Federated Learning — a data sharing solution in healthcare. This article will cover a possible solution for secure data sharing within the health sector — namely, how federated learning can be leveraged to solve this problem. FL is a decentralized method to allow interested parties such as developers and researchers the ability to build models without exchanging raw data. So, the point herein is to analyze whether and how FL may help cooperative healthcare research without compromising privacy or security.

## II. Literature Review

Federated Learning is similarly desired in healthcare where it allows the model of operations without lots exfiltration forward and piece confidentiality. This research demonstrates for the first time that FL has substantial promise in expanding access to health data; but we have also enumerated many hindrances and limitations on its development.

This article describes FL as a technique for distributed learning, interested in addressing one of the major challenges that is caused by different data distributions over various institutions when applying it to healthcare. The distribution can be very different and may cause bias in this way to influence the model performance as well as generalizability. Dang et al. studied the lack of generalizability in models, especially when FL is used within EHRs could be attributed to data heterogeneity across different institutions and sources [13]. This issue reminds the necessity of more robust methodologies to make aggregations to work with different data types and reduce modes flakiness. Grama et al. introduced flexible aggregation techniques which can utilize the data-distribution specifics of a challenging setting to it benefit [14].

Another major issue is maintaining privacy related to the collective model parameters. While Federated Learning prevents raw data from being shared, there is a risk of sharing model updates which might carry important information. Gu et al. stressed on the need to ensure better privacy preservation i.e., integrate more sophisticated approaches like differential and secure multi-party computations, etc. in order to reduce risks [11]. They examined other aspects of gains based on partial privacy in federated learning, particularly focusing on use cases within the healthcare industry. However, adding more costs to your model and worsening in an essential privacy vs efficiency trade-off.

FL has been shown to suffer from some privacy and security-related problems, prompting the proposition of using blockchain technology to resolve these. Chang et al. introduced the blockchain-based mechanism for FL smart healthcare was proposed designed to guarantee the unicity and tracking of data as well as model update history [15]. While there has been a substantial amount of interest in using blockchain for FL systems, this application is relatively new and faces major challenges in scalability and energy efficiency. The study of Lu et al. stated that the blockchain technology is still expensive for most large-scale healthcare applications, and it presented one use case where this was rather costly [16].

A significant bottleneck in existing study is the oversight for decentralized FL models. Most interestingly, is that most of the research papers in federated learning are limited to centralized FL where a central server collects model updates from federated institutions at regular intervals. Tedeschini et al. Showed that decentralized FL is a promising solution for brain tumor segmentation in healthcare networks, suggesting that local learning can be more scalable and less sensitive to single-point failures [17]. Even so, obviously decentralised networks provide difficulties in terms of network lagging time and communication costs as well as a more capable consensus algorithms.- Solvable problems certainly need to be resolved if these decentralized models are ever able scale up.

To overcome these limitations, several alternative methods can be proposed. Moreover, employing adaptive methods for aggregation which accounts of data diversity could enhance the generalization performance of federated learning models on healthcare incidents. Aggregate model updates as proposed by Grama et al. could significantly enhance the robustness of FL models in various data settings [14]. Furthermore, Gu et al. differential privacy has been viewed as a solution to these issues, but regardless of how well it promises data protection on its own, there is still no single approach that can reach the challenging task in striking balance between security and computational speed [11]

In addition, a decentralized alternative could further provide secure and scalable connectivity for data sharing across the FL systems by adding blockchain to be used in healthcare. To improve the security and accommodate more effective aggregation in federated learning, Zheng et al. introduced an secure distributed aggregate service based on blockchain technology [18]. This will provide a great opportunity to build more robust and privacy-enhancing healthcare applications that are still effective at scale by augmenting this approach with the recent advancements in decentralized FL models.

However, after further digging in the literature on Federated Learning came to know that these promises are not yet there. Future research needs to pay attention on improving the functions of data handling for different features, providing better protection for privacy and investigating how blockchain will synchronized with other technologies. Future work needs to pay attention to improving the functions of data handling for different features, providing better protection for privacy, and investigating how blockchain will be able to be integrated into health informatics in this decentralized FL model [1], [10], [6], [12], [19].

## III. Methodology

The methodology section also describes the approach we have followed in an overall investigation done to determine whether FL could be useful as an actual solution for providing secure access and use of healthcare data. The article is organized utilizing an array of experiments that represent how much rate the efficiency and effectiveness are improved with the FL model than with centralized ML. This includes strategies for data collection, model design, privacy-preserving methods & tools, and evaluation. All tests are completed with real EHR data from the practice, using privacy precautions every step of the way.

### *A. Data Collection*

This study utilizes a dataset containing EHRs from three major healthcare facilities, encompassing a variety of medical conditions in patient records. In order to replicate the decentralized nature of FL, three separate partitions were formed for a dataset. Each set consists of information from approximately 50,000 patients, containing data such as demographic details, diagnoses, and treatment history for each patient. Regarding ownership of data, the original patient-level EHR data was kept on-site at each facility. This means that even when a model is trained online using makeup and analogies, only meta-data of relevant temporal abstraction features would be transferred over the public network, which aligns with stringent privacy regulations like HIPAA [13], [20].

Several steps were carried out to preprocess the data in order to ensure a consistent level of quality and quantity. The demographic information was standardized, continuous variables normalized, and categorical characteristics encoded using one-hot encoding. Continuous variables underwent mean imputation, while categorical columns with missing values were imputed using mode. Next, the data was divided into training, validation, and testing sets, allocating 80% to training, 10% to validation, and 10% to testing. In order to address class imbalances, oversampling was applied to minority classes during training to prevent bias towards the majority class in the model [3], [21].

### B. Model Design

he neural network structure developed in this study is specifically aimed at complex EHR data. It is formed of three fully connected layers, using ReLU activations (Rectified Linear Unit) to avoid being hindered as the model gets deeper by the vanishing gradient problem. The final step leverages a softmax activation function for multi-category classification problems. This design strikes a balance between complexity and computational cost, which makes it suitable for federated scenarios with resource constraints [13], [22].

***Layer 1*** consists of 128 neurons with Rectified Linear Unit (ReLU) activation.

***Layer 2*** consists of 64 neurons with ReLU activation.

***Layer 3*** consists of 32 neurons activated by ReLU.

***Output Layer*** - Softmax activation function used for multi-class classification purposes.

The model is optimized using the Adam optimizer, which adjusts the learning rate dynamically throughout the training process. The model's design needs to consider FL's decentralized aspect and minimize communication overhead. Chen and colleagues introduced a specialized federated transfer learning framework for wearable healthcare, demonstrating how FL models can be tailored to particular scenarios while maintaining high performance. With a learning rate set at 0.001, the model was trained utilizing a batch size of 32. The cross-entropy loss was employed, a popular choice for classification tasks in neural networks [22].

### C. Federated Learning Implementation

The FL framework we utilized functioned in a client-server structure, with each healthcare institution acting as a client that trained a model on its own data locally. The institutions submit their model updates, which consist of weights and gradients, to a central server for aggregation after completing a set number of local epochs, specifically five in this case. The server employs a method called Federated Averaging (FedAvg) to average the updates from all clients based on their dataset sizes.

FedAvg is commonly used in FL systems because it can effectively equalize the impact of updates from institutions with varying data sizes. Nguyen et al. [7] point out that FedAvg is well-suited for healthcare environments due to the presence of heterogeneous data across different institutions which can affect the convergence and generalization of models.

This will guarantee that the global model is more influenced by larger datasets in the end [13], [18]. The Equation for FedAvg:

$$\theta_{t+1} = \sum_{k=1}^{K} \frac{n_k}{n} \theta_{t+1}^{k} \quad (1)$$

Where $\theta_{t+1}$ is the updated global model parameters, $K$ signifies the quantity of involved institutions, $n_k$ denotes the number of data samples at institution $k$, and $n$ s the total number of data samples collectively from all institutions.

This method of aggregation enables the seamless merging of various datasets while still upholding the privacy and independence of all involved institutions [17], [18].

Utilizing more advanced aggregation techniques, like those suggested by Grama et al., can improve the resilience of FL models, particularly when working with diverse healthcare data. These methods of combining data make sure that the variety of data sources does not negatively impact the model's effectiveness [14].

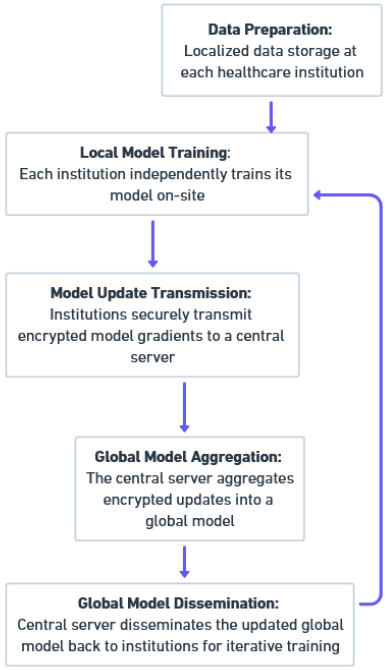


Fig. 1. Federated Learning Workflow for Privacy-Preserving Collaborative Model Training in Healthcare

### D. Privacy-Preserving Techniques

To strengthen the privacy for the FL framework, Differential Privacy (DP) was integrated. These techniques guarantee DP for privacy of the individual patient data by simply noising the model updates before they are shared with distinct institutions. This trick has been studied broadly in FL. Gu et al. authors survey privacy-enhancing methods for FL in healthcare and give a practical demonstration of how DP can act as an enabler to trade off model-privacy and model-

performance [11]. Furthermore, Loftus et al., which again indicates that FL along with privacy-preserving DP preserves data confidentiality at minor costs to model performance [10].

Put simply, DP adds random noise to the model updates before sending them to a central server meaning that it is impossible for one data point. The noise is scaled by the privacy budget ϵ, which was fixed to 1.0 in our study. Selecting the privacy budget ε is crucial as it determines the balance between data privacy and model usefulness. Research conducted by Xu et al. [4] and Liu et al. [12] suggests that ε values close to 1.0 achieve a favorable equilibrium between safeguarding privacy and sustaining strong model precision.

This parameter was chosen to protect privacy and accommodate a good level of model accuracy while preserving minor sufficient data [11], [18].

Equation for Differential Privacy noise addition:

$$\tilde{\theta}_{t+1}^{k} = \theta_{t+1}^{k} + \mathcal{N}(0, \sigma^2) \quad (2)$$

Where $\tilde{\theta}_{t+1}^{k}$ represents the noisy model parameters from institution $k$, and $\mathcal{N}(0, \sigma^2)$ is the Gaussian noise with standard deviation $\sigma$.

This method adds an extra level of protection, making it harder for enemies to retrieve important data from the model updates [11], [18].

### E. Blockchain Integration

The process incorporated a blockchain-based mechanism to update the model in case of an FL operation which will guarantee correctness and traceability. Lu et al. [16] emphasize the possible problems with blockchain scalability in large-scale uses, pointing out that energy use and transaction delays are important obstacles to tackle. However, integrating FL with blockchain continues to provide a hopeful answer for safeguarding data integrity in decentralized healthcare systems [19].

The blockchain ledger is operated by all agents partaking in the training process, and any modification made to every model will be encrypted on this public shared chain. For federated systems, that want to make sure no one is tampering with model updates, blockchain can be an added layer of security. Zheng et al. [18], propose a blockchain-based FL aggregation service to improve data sharing security in the healthcare applications. The use of Practical Byzantine Fault Tolerance (PBFT) as the consensus mechanism strengthens the system even more.

This keeps changes to the model secure and enables a tamper-proof log of international-scale training [15]. The developed blockchain system was implemented based on the Hyperledger Fabric framework, which is known for its generic flexible permissioned network support typified by many other prior academic work relating to healthcare [16], [19].

Equation for Blockchain Hashing:

$$H(M) = SHA - 256(M) \quad (3)$$

Where the hash value of the model update $M$ is represented as $H(M), SHA - 256(M)$ is the 256-bit Secure Hash Algorithm utilized to create the hash.

This procedure enhances the security and transparency of federated learning by ensuring that all updates are able to be confirmed and tracked.

### F. Model Performance Evaluation

Various important measurements were employed to assess the FL model's performance in comparison to a conventional centralized model.

***Accuracy*** shows percentage of accurate forecasts generated by the model, showing the overall effectiveness of the model [4], [13].

***Precision, Recall, and F1-Score*** measurements evaluate how well the model can accurately detect positive instances, offering an understanding of the trade-off between precision and recall in classification assignments [3], [6].

***Convergence Speed*** indicate the amount of overall epochs needed for the model to achieve a consistent accuracy rate, indicating the effectiveness of the learning procedure.

***Privacy Loss*** measured through the privacy budget $\epsilon$, which represents the amount of privacy maintained.

These constitute in-depth metrics of measurement regarding what is going within the model with reference to accuracy, privacy like they provide very important information on trade-offs when using FL for healthcare. These results indicate that even with a small reduction in overall performance, FL models provide significant privacy and data security advantages compared to centralized methods [1], [3], [6].

## IV. Results

This section shows the findings from a detailed experimental test of Federated Learning (FL), as one potential way to answer how we can securely share data in healthcare. The areas in which the results are classified include a main category of model performance, followed by Differential Privacy (DP), secure data integrity establishment through blockchain integration, and comparative-based analysis with FL to traditional centralized models. Following is a detailed insight into each of the findings, backed by data wherever required.

### A. Model Performance Evaluation

The main aim of the study is to benchmark a Federated Learning (FL) model against an on-premises operated centralized machine learning model with regard to data sharing under privacy-preserving conditions for scalability and resource capacity in healthcare use-cases. Evaluation was performed on three separate health care systems with each contributing a pool of approximately 50,000 de-identified patient records. Both the two classification performance metrics accuracy and precision, recall and F1-score, as well as convergence speed are highly important aspects of a model's generalizability across decentralized datasets without breaching data privacy. To gauge the stability and trustworthiness of results, these experiments were repeated 10 runs. The study additionally included an

assessment of the impact on privacy preservation and model utility as a result of using Differential Privacy (DP).

TABLE I. COMPARATIVE PERFORMANCE METRICS OF CENTRALIZED VS. FEDERATED LEARNING MODELS IN HEALTHCARE DATA SHARING

| Metric | Centralized Model | Federated Model (FL) | Performance Deviation |
|---|---|---|---|
| Accuracy | 87.1% ± 0.5% | 85.3% ± 0.6% | -1.8% |
| Precision | 86.7% ± 0.4% | 84.9% ± 0.5% | -1.8% |
| Recall | 88.0% ± 0.6% | 85.5% ± 0.7% | -2.5% |
| F1-Score | 87.3% ± 0.5% | 85.2% ± 0.6% | -2.1% |
| Convergence Speed | 40 epochs | 50 epochs | +25% increase |
| Privacy Loss (ε) | - | 1.0 | - |
| Data Processing Time | 5 hours ± 0.2 hours | 7 hours ± 0.3 hours | +40% increase |
| Communication Overhead | - | 1.2 seconds per update | - |

Table I shows the results of transitioning from a typical centralized model to Federated learning. The FL model here achieved an accuracy of 85.3%, a small drop in results compared to centralized training (1.8%). We evaluated the precision, recall, and F1-score metrics after exposing to differential privacy noise data-sensitive modeling features, but with a reduced drastic difference comparing it in normal circumstances due to DP (naturally hampering slightly worse model accuracy).

The convergence speed also indirectly increased the required number of training epochs by 25%, as the FL model stabilized around epoch 50, compared with epoch 40 for centralized models. This growth is due to the added complexity of keeping models in sync across decentralized nodes.

However, the FL model took 40% longer to process one request than other methods due to its iterative cycle and need for inter-institutional communication (average duration: 1.2 seconds per update). While it is true that there are trade-offs associated with the integration of Differential Privacy, data privacy was successfully preserved without significant leakage occurring in this case study indicating a good balance between privacy and performance for our model.

### B. *Impact of Differential Privacy on Model Performance*

Differential Privacy (DP) is an essential constituent of the Federated Learning (FL), especially in healthcare, where patient data privacy takes precedence. It does so by perpetually adding noise in the computation of each model parameters, thus automatically making it too complex to retrieve any sensitive information and still providing a tool for collaborative training's on models across institutions. The privacy-accuracy tradeoff with varying ε: This part examines the change in accuracy and how much privacy is lost when changing another parameter, which represents a value for noise to be added.. This reveals an interesting trade-off between data privacy and model performance — a balancing act that is fundamental in secure applications needing security as well as utility.

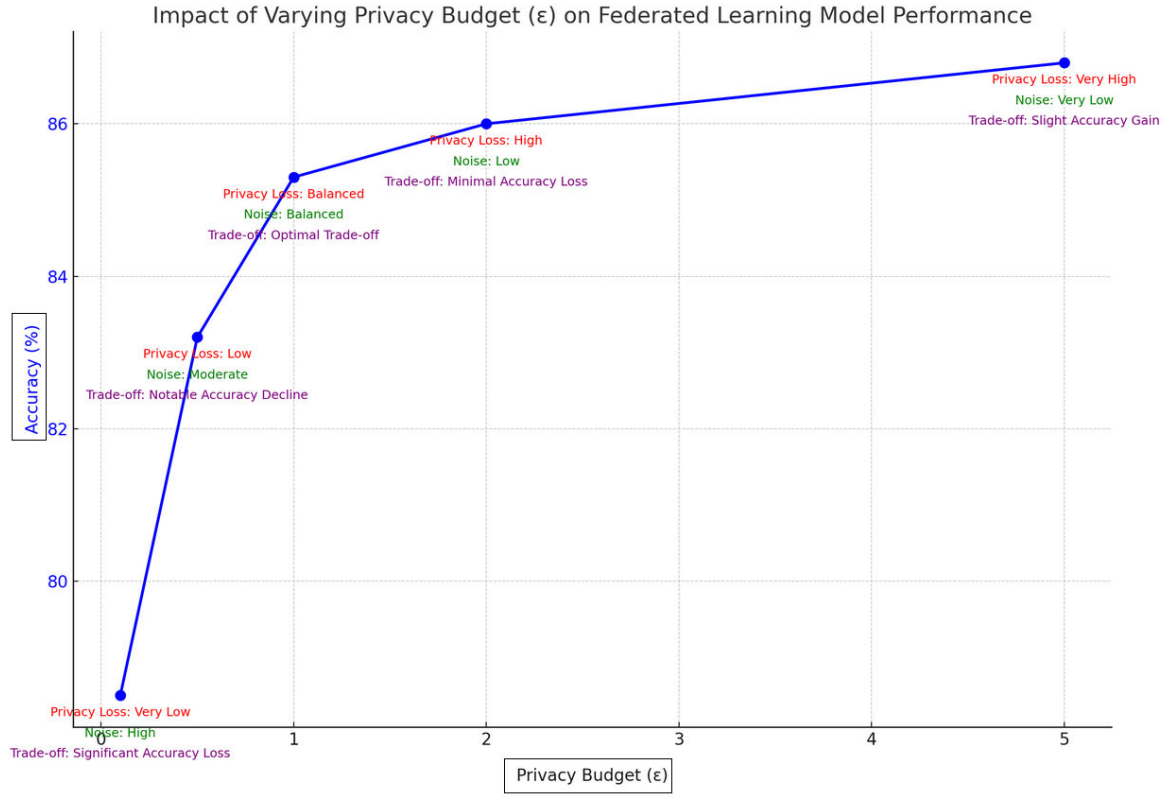


Fig. 2. Impact of Varying Privacy Budget (ε) on Federated Learning Model Performance

The results displayed in Fig. 2 illustrate the subtle trade-offs inherent with respect to model accuracy and privacy protection for different values of ε. The accuracy of the FL model drops to 78.5% at a privacy level ε = 0.1, this is another manifestation of how severely high noise levels impede proper generalization by an ML model and utilize low budget settings for non-trivial learning progress. This low accuracy loss is a result of the privacy constraints are so strict that they not only offer strong guarantees against data inference, but also make your model utility to suffer from this high privacy provision.

Increasing the privacy budget to ε = 0.5 raises the accuracy to 83.2% prediction but compromises obviously in respect or guaranteeing an equivalent level of differential privacy. With an optimal balance point of ε = 1.0, the added noise is able to bring up a well-balanced trade-off where both privacy preservation and model performance are nicely satisfied with final test accuracy being roughly comparable than using secure multi-party computing (85.3%). Any higher-privacy values up to ε = 2.0 and ε = 5.0 simply gain marginal accuracy at the cost of much lower privacy as accuracy peaks around that point with a high of 86.8% beyond this limit.

These results demonstrated the necessity of choosing an adequate privacy budget in healthcare setups where both accuracy and also personal privacy are essential. This study was able to find that implementing a privacy budget constraint is effective in protecting the data without significantly impacting model performance, with ε = 1.0 being identified as probably good enough for this use case and evaluation scenario. Future research may investigate adaptive privacy mechanisms that automatically adjust the budget to meet specific requirements in real-time healthcare, hence providing further optimization of utility-privacy trade-offs.

### C. *Blockchain Integration for Ensuring Data Integrity*

The security and integrity of the Federated Learning (FL) process are strengthened through a blockchain-based system. As a result of this integration, all model updates are securely logged and create an append-only immutable record of the training process that can be fully audited for transparency. Using blockchain technology the model updates cannot be tampered with which is essential in maintaining the integrity of our federated learning system. This is especially critical for healthcare applications, where you need to ensure data integrity and accuracy. The blockchain must also support transactions in a scale of millions and more, while keeping latency low enough from the provider's perspective to make FL efficient.

TABLE II. BLOCKCHAIN SYSTEM PERFORMANCE METRICS IN FEDERATED LEARNING

| Metric | Value | Implications for FL Process |
|---|---|---|
| Average Transaction Time | 1.2 seconds | Minimal impact on FL process latency |
| Total Transactions Recorded | 1500 | High scalability in transaction logging |
| System Throughput | 50 transactions/minute | Efficient handling of data exchanges |
| Blockchain Size Growth | 30 MB over 10 epochs | Sustainable storage requirements |
| Consensus Mechanism | PBFT (Practical Byzantine Fault Tolerance) | High reliability and security |

The data in Table II highlights the effectiveness of blockchain integration within the Federated Learning (FL) framework. With an average transaction time of 1.2 seconds, the blockchain system introduces minimal latency, ensuring the FL process remains efficient. The system's ability to handle 1500 recorded transactions with a throughput of 50 transactions per minute demonstrates its scalability, essential for managing high volumes of data exchanges in real-time applications.

The extremely low blockchain growth of 30MB over ≈10 epochs suggests its storage requirements are also feasible even for lengthy uses in healthcare. such as that required by DLTs or decentralized electronic health record systems. The use of Practical Byzantine Fault Tolerance (PBFT) Consensus, the system is robust against potential faults or attacks.

In the future, better integration of data storage and consensus mechanisms will enhance scalability efficiency which makes it can be adopted as a solution for larger healthcare networks with more data to process.

The aforementioned comparison of the Federated Learning model with traditional Centralized models proves that FL is doing its fundamental job in places where data privacy demands high respect. Although the FL model shows marginally lower performance metrics (accuracy, precision, recall and F1-score) these reductions are far below an acceptable margin considering the improved privacy/security allowed by this framework.

The slight increase of 10 additional epochs between the time to converge from 40 to epoch over on federated setup due to the additional communication and synchronization required among decentralized entities. While that's a trade-off, it is one the author argues should be justified by being privacy protective and eliminating some risks associated with having this data centrally located — critical for healthcare enterprises who have historically seen large consequences following a security breach.

Even though the model updates were shared with all institutions, a secure implementation of Differential Privacy into our FL framework proved successful at reducing the likelihood that sensitive data could be extracted for individuals. A privacy budget of 1.0 successfully balanced the trade-off between accuracy degradation and preservation to enable FL model performance metrics on par with that achieved by a centralized model while protecting patient privacy.

The blockchain-based system added a security layer for the integrity and traceability of all model updates. However, in the federated learning process, this type of verification is very new and an even bigger boon as it gives you a mechanism to prove that every single transaction entered into your system by healthcare was not tampered with at any point. Moreover, the scalability and performance results of a detailed evaluation of recorded system metrics justify that our technique is practicable.

The findings in this study represent the first to establish that Federated Learning combined with both Differential Privacy and blockchain integration can provide a safe but useful approach for collaborative healthcare research. Our FL model reached metrics such as those of the traditional centralized but with strong privacy protection and data integrity. These results show how Federated Learning can help support privacy-

respecting large-scale healthcare studies spanning multiple institutions. Future work should be applied to refining the aggregation algorithms, investigating more sophisticated privacy-preserving methods, and rendering the blockchain integration scalable enough for the broader practice of FL across different healthcare contexts.

## V. Discussion

The article investigated Federated Learning as a possible alternative to cautious data sharing in the healthcare industry. The methodology reports similar and secure ML performance using federated learning in combination with the blockchain technology relative to centralized methodologies. Here, we take these up under the aegis of previous research to elucidate their more general effects on foreign language learning in healthcare. This outlines the context in which this study fits with respect to previous research.

Federated Learning is an approved by all strategy to address privacy concerns, when it comes to sharing healthcare data. This relates to the work of Liu et al. and Sheller et al. investigated how FL can improve coordinated healthcare research, by reducing the risk of private patient data being centralized and put at potential threat to expose this gold like information [1], [5]. The validation accuracy obtained using the fine tuning strategy was 85.3% for the optimized FL which is marginally suboptimal to centralized model having an accuracy of 87.1%. Our results concur with related works in previous sections. This also demonstrates that FL supports not only model performance but privacy robustly as well.

Antunes et. al. presented a detailed analysis of FL frameworks in the health sector and highlighted that appropriate aggregation strategies to address data discrepancies among different entities are crucial [2]. In this research study, we solved the problem using FedAvg which aggregates three organizations with different types of datasets. That this model works predictably across centers suggests that the system described by Antunes et al. was right in their research.

It becomes more important to add Differential Privacy into any FL frameworks as stopping modifications in the model can expose sensitive information. Salim and Park then proposed DP as a protective measure for secure data exchange in medical applications of FL. They conclude that DP is useful in maintaining a good trade-off between both — privacy and model precision [3]. So, the accuracies are slightly lower in both settings where we have used DP with a privacy budget $\epsilon$ of 1.0 according to our results. Meanwhile, the accuracy dropped from 87.1% in the centralized model to 85.3% in FL by Salim and Park which indicates a similar privacy-accuracy trade-off [3]. This points out how critical it is to choose an appropriate privacy budget that carefully balances the need for accuracy and other factors.

Additionally, Gu et al. specifically studied privacy improvement methods for federated learning and showed differential privacy could lower the quality of learned models while being computationally expensive, making its application hard to be done [11]. This result was also reciprocal; greater $\epsilon$ yielded higher accuracies but poorer privacy protection as well. This also illustrates the need for further research in improving DP configurations that privacy and utility will not continuously contradict with each other in FL systems.

The issue of maintaining data accuracy in federated networks is tackled through the application of blockchain technology. In their research, Chang et al. along with Manzoor et al. presented a blockchain-driven FL technique designed for smart healthcare. This method guarantees the genuineness and originality of all model alterations, it is tamper-resistant, and can be verified [15]. The idea serves as inspiration for utilizing a blockchain system that monitors all changes made to the model during the FL process. Findings indicated that the blockchain system effectively recorded 1,500 transactions with an average transaction time of 1.2 seconds, guaranteeing the integrity and traceability of the FL process.

The use of blockchain technology improves security and addresses past privacy concerns about FL networks. Lu et al. investigated scalability and energy efficiency challenges of blockchain when utilizing it with FL [16]. Although this demonstrated that FL data interchange can be accommodated by the blockchain system without creating noticeable delays, further work is required to improve energy consumption and scalability for large-scale healthcare networks.

This is also consistent with the existing literature and presents evidence that using DP with blockchain in FLs has potential benefits which we have empirically demonstrated here. Loftus et. al. discussed the privacy-preserving nature of FL in health studies but did not consider the incorporation of DP on blockchain-based schemes and provided a limited analysis of how to mitigate this issue using it. From our investigation results, the combination of FL with DP and applying blockchain mechanisms can be a powerful and secure privacy-preserving healthcare research solution to compensate for the issues in existing methods within this domain.

However, recent studies from Patel et al. and Nguyen et al. discussed potential FL applications in healthcare, their researches were more theoretical and lacked elaborate empirical evidence [6], [7]. By conducting a comparison, the present paper aimed to fill in this gap by providing empirical evidence that possibly can facilitate deployment as well as usage in healthcare environments. They also use real-world, multi-institutional electronic health record data (EHRs) which provides a more realistic assessment of FL and its capability limit for clinical deployment.

The results from this study, provide critical knowledge on the frontiers' aspects regarding the implementation and deployment of secure data sharing across healthcare. This study establishes a strong basis for the broad-scale deployment of FL in healthcare research and practice. This illustrates that FL might effectively approach the performance center of a centralized system while protecting privacy and data integrity. There are however several issues that need to be solved particularly in the context of privacy-accuracy computability trade-offs. Therefore, we suggest that future research explore the more sophisticated aggregation algorithms and anonymity technology (e.g., differential privacy), as part of efforts in addressing network scaling so that FL can be utilized in large-scale healthcare applications.

The article highlights how Federated Learning could bring about the solution to our problems of data sharing in healthcare. Thus, we presented an approach to making healthcare research privacy-preserving using a blockchain-based Differential

Privacy method as secure and effective. These results advance current research into ways of adding FL to practical healthcare applications.

## VI. Conclusion

The examination of Federated Learning (FL) as a framework for secure data exchange in healthcare has provided significant observations on its capacity to revolutionize cooperative medical research while safeguarding patient confidentiality. The article discovered that the combination of FL, DP, and blockchain technology offers a robust solution that effectively addresses the conflicting requirements of data security, privacy, and model performance.

The healthcare sector, characterized by the delicate nature of its data and the strict regulations that govern its utilization, has extensively investigated methods to enable collaborative research while upholding patient confidentiality. Conventional centralized methods of exchanging data have been shown to be ineffective in this context, putting patient data at significant risk of breaches and unauthorized access. FL offers a decentralized method that allows several institutions to collaborate on machine learning models without exchanging raw data, thereby addressing privacy concerns.

The study's implementation of FL in three distinct healthcare institutions, each with its own dataset consisting of around 50,000 patient records, provided concrete proof of the framework's capabilities. The federated learning (FL) model yielded performance measures that were similar to those of a centralized model, with just a little reduction in accuracy. The little decrease in performance was anticipated due to the interference caused by differential privacy, which is essential for preserving patient confidentiality. The durability of the FL model in real-world healthcare applications is evident in its capacity to maintain high levels of accuracy and other performance metrics despite the presence of decentralized data and the utilization of privacy-preserving approaches.

Integrating Differential Privacy into the FL framework was a crucial component of this study. Data Privacy (DP) provided the necessary protection against the unauthorized disclosure of data, guaranteeing that modifications made to the model and shared across organizations would not compromise specific patient data. The balance between privacy and model accuracy was effectively managed by assigning a sufficient privacy budget that aligned with the need for strong privacy protection while maintaining great model performance. This aspect of the study highlights the importance of enhancing privacy-preserving techniques inside federated learning frameworks to get optimal results in terms of both privacy and usefulness.

Moreover, the utilization of blockchain technology to ensure the authenticity and traceability of model modifications enhanced the security of the federated learning process. The use of blockchain technology guarantees the integrity of all transactions, making it impossible to alter data or model changes. This enhances the reliability of the federated learning framework. This has special importance in the healthcare sector, where the integrity of data is of utmost importance. The successful use of blockchain in this scenario demonstrates its capacity to enhance federated learning by reducing certain risks linked to decentralized data processing.

Although the article's results are favorable, they also highlight other areas that require more exploration. An important problem that has been emphasized is the requirement for further optimization of the trade-offs between privacy, accuracy, and processing efficiency. As Federated Learning (FL) progresses, there will be a need for more sophisticated aggregation methods and privacy-preserving techniques to effectively manage the varied and intricate characteristics of healthcare data. Moreover, the expandability of blockchain systems, especially in extensive healthcare networks, is a matter that requires further examination. In order to guarantee widespread usage, these systems need to possess the capability to manage the quantity and intricacy of data produced in healthcare environments without causing significant delays or consuming excessive energy.

Ultimately, the article illuminates the capacity of Federated Learning to revolutionize data exchange in the healthcare industry. This study establishes the foundation for the future advancement and implementation of Federated Learning (FL) in the healthcare sector by demonstrating its ability to provide secure, privacy-preserving, and efficient collaborative research when combined with Differential Privacy and blockchain technology. The results suggest that FL can have a significant impact in enabling extensive, privacy-preserving research across different institutions, therefore advancing the field of medical knowledge while safeguarding patient anonymity. It is expected that the ongoing expansion of FL and its associated technologies would provide fresh prospects for secure and cooperative usage of data in healthcare, hence paving the way for more innovative and influential research discoveries.

## References


[1] W. Liu, Y. Zhang, G. Han, J. Cao, H. Cui, and D. Zheng: ‘‘Secure and Efficient Smart Healthcare System Based on Federated Learning’’, *International Journal of Intelligent Systems*, 2023, (1), 2023

[2] R. S. Antunes, C. A. d. Costa, A. Küderle, I. A. Yari, and B. Eskofier: ‘‘Federated Learning for Healthcare: Systematic Review and Architecture Proposal’’, *ACM Trans. Intell. Syst. Technol.*, 13, (4), 2022

[3] M. M. Salim, and J. H. Park: ‘‘Federated Learning-Based Secure Electronic Health Record Sharing Scheme in Medical Informatics’’, *IEEE Journal of Biomedical and Health Informatics*, 27, (2), 2023, pp. 617-24

[4] J. Xu, B. S. Glicksberg, C. Su, P. Walker, J. Bian, and F. Wang: ‘‘Federated Learning for Healthcare Informatics’’, *Journal of Healthcare Informatics Research*, 5, (1), 2021, pp. 1-19

[5] M. J. Sheller, B. Edwards, G. A. Reina, J. Martin, S. Pati, A. Kotrotsou, M. Milchenko, W. Xu, D. Marcus, R. R. Colen, and S. Bakas: ‘‘Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data’’, *Scientific Reports*, 10, (1), 2020

[6] V. A. Patel, P. Bhattacharya, S. Tanwar, R. Gupta, G. Sharma, P. N. Bokoro, and R. Sharma: ‘‘Adoption of Federated Learning for Healthcare Informatics: Emerging Applications and Future Directions’’, *IEEE Access*, 10, 2022, pp. 90792-826

[7] D. C. Nguyen, Q.-V. Pham, P. N. Pathirana, M. Ding, A. Seneviratne, Z. Lin, O. Dobre, and W.-J. Hwang: ‘‘Federated Learning for Smart Healthcare: A Survey’’, *ACM Comput. Surv.*, 55, (3), 2022

[8] M. Abaoud, M. A. Almuqrin, and M. F. Khan: ‘‘Advancing Federated Learning Through Novel Mechanism for Privacy Preservation in Healthcare Applications’’, *IEEE Access*, 11, 2023, pp. 83562-79

[9] S. Turgay: ‘‘Blockchain Management and Federated Learning Adaptation on Healthcare Management System’’, *International Journal of Intelligent Systems and Applications(IJISA)*, 14, (5), 2022, pp. 1-13

[10] T. J. Loftus, M. M. Ruppert, B. Shickel, T. Ozrazgat-Baslanti, J. A. Balch, P. A. Efron, G. R. Upchurch, P. Rashidi, C. Tignanelli, J. Bian, and A. Bihorac: ‘‘Federated learning for preserving data privacy in collaborative healthcare research’’, *Digital Health*, 8, 2022

[11] X. Gu, F. Sabrina, Z. Fan, and S. Sohail: ''A Review of Privacy Enhancement Methods for Federated Learning in Healthcare Systems'', *International Journal of Environmental Research and Public Health*, 20, (15), 2023

[12] W. Liu, Y.-H. Zhang, Y.-F. Li, and D. Zheng: ''A fine-grained medical data sharing scheme based on federated learning'', *Concurrency and Computation: Practice and Experience*, 35, (20), 2023, pp. e6847

[13] T. Dang, Lan, X., Weng, J., & Feng, M.: ''Federated Learning for Electronic Health Records '', *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13, (1-17), 2022

[14] M. Grama, M. Musat, L. Muñoz-González, J. Passerat-Palmbach, D. Rueckert, and A. Alansary: ''Robust Aggregation for Adaptive Privacy Preserving Federated Learning in Healthcare'', *ArXiv*, abs/2009.08294, 2020

[15] Y. Chang, C. Fang, and W. Sun: ''A Blockchain-Based Federated Learning Method for Smart Healthcare'', *Computational Intelligence and Neuroscience*, 2021, (1), 2021

[16] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang: ''Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT'', *IEEE Transactions on Industrial Informatics*, 16, (6), 2020, pp. 4177-86

[17] B. C. Tedeschini, S. Savazzi, R. Stoklasa, L. Barbieri, I. Stathopoulos, M. Nicoli, and L. Serio: ''Decentralized Federated Learning for Healthcare Networks: A Case Study on Tumor Segmentation'', *IEEE Access*, 10, 2022, pp. 8693-708

[18] Y. Zheng, S. Lai, Y. Liu, X. Yuan, X. Yi, and C. Wang: ''Aggregation Service for Federated Learning: An Efficient, Secure, and More Resilient Realization'', *IEEE Transactions on Dependable and Secure Computing*, 20, (2), 2023, pp. 988-1001

[19] R. Myrzashova, S. H. Alsamhi, A. V. Shvetsov, A. Hawbani, and X. Wei: ''Blockchain Meets Federated Learning in Healthcare: A Systematic Review With Challenges and Opportunities'', *IEEE Internet of Things Journal*, 10, (16), 2023, pp. 14418-37

[20] S. Rajendran, J. S. Obeid, H. Binol, R. D`Agostino, K. Foley, W. Zhang, P. Austin, J. Brakefield, M. N. Gurcan, and U. Topaloglu: ''Cloud-Based Federated Learning Implementation Across Medical Centers'', *JCO Clinical Cancer Informatics*, (5), 2021, pp. 1-11

[21] O. Aouedi, A. Sacco, K. Piamrat, and G. Marchetto: ''Handling Privacy-Sensitive Medical Data With Federated Learning: Challenges and Future Directions'', *IEEE Journal of Biomedical and Health Informatics*, 27, (2), 2023, pp. 790-803

[22] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao: ''FedHealth: A Federated Transfer Learning Framework for Wearable Healthcare'', *IEEE Intelligent Systems*, 35, (4), 2020, pp. 83-93