International Journal of Scientific and Management Research
Volume 8 Issue 8 (August) 2025
ISSN: 2581-6888
Page: 52-72

A Federated Learning Approach to Secure AI-Based Patient Outcome Prediction Across Hospitals

Sarah Mavire[1]*, Kumbirai Bernard Muhwati[2], Carrol Donna Kudaro[3], & Joy Awoleye[4]
[1]Department of Computer Science, Yeshiva University, **USA**
[2]Department of Computer Science, Yeshiva University, **USA**
[3]Department of Computer Science, Yeshiva University, **USA**
[4]Department of Computer Science, Yeshiva University, **USA**

DOI - http://doi.org/10.37502/IJSMR.2025.8806

Abstract

The potential of artificial intelligence to transform healthcare is increasingly realized through patient outcome prediction models. However, traditional centralized training methods for such models pose significant privacy risks, particularly when sensitive patient data must be shared across institutions. This paper proposes a federated learning (FL) framework for developing robust and secure patient outcome prediction models across hospitals while ensuring data privacy and regulatory compliance. Using synthetic and real-world datasets such as MIMIC-III and eICU, it creates a multi-hospital-based environment, where models are trained locally and then aggregated in a centralized manner without sharing raw patient data. The LSTM-based and transformer-based architectures are being applied in experiments to time-series health record data, and the accuracy of prediction is statistically significant in the outcomes of ICU mortality and readmission. The FL model achieves competitive performance compared to centralized training, with less than 3% performance degradation and full compliance with privacy-preserving standards. Differential privacy and secure aggregation enhancements was also explored to improve robustness against adversarial participants. Our findings indicate that federated learning presents a scalable, secure, and practical approach to collaborative AI in healthcare, bridging the gap between innovation and privacy protection.

Keywords: Federated Learning (FL), Artificial Intelligence (AI), patient outcome prediction, healthcare, privacy, Electronic Health Records (EHRs), Differential Privacy (DP), Secure Aggregation, LSTM, transformer-based architectures, MIMIC-III, eICU.

1. Introduction

In recent years, the application of artificial intelligence (AI) in the health care sector has grown significantly, offering revolutionary features for patient treatment and clinical decision-making. AI models have proven to be extremely proficient in predicting patient outcomes such as in-hospital mortality, disease progression, readmission to hospital, and response to therapy (Rajkomar et al., 2018). Machine learning (ML) and deep learning (DL) algorithms are increasingly being incorporated into clinical systems and help clinicians to diagnose chronic diseases, customize treatment protocols, and forecast pivotal events like sepsis or heart failure (Esteva et al., 2019). These advances have been facilitated by the availability of large electronic health records (EHRs), powerful computing capabilities, and increasingly sophisticated neural network models such as recurrent neural networks (RNNs) and transformers. In intensive care units (ICUs), machine-learning models trained on databases such as MIMIC-III and eICU have helped forecast mortality and ventilator dependence (Johnson et al., 2016). Similarly, in chronic disease management, AI models allow personalized treatment guidelines based on patient history and lab findings. But to achieve the maximum potential of AI, the models need to be granted access to large and diverse datasets that accurately reflect patient populations across a series of various institutions. This requirement gives rise to serious concerns over data ownership, privacy, and control.

Although clinically useful, centralized training of AI carries substantial risks and limitations. The majority of the conventional AI models are trained on data collected and compiled from a single or very few institutions, raising issues regarding the model's generalizability and fairness. In addition, condensing sensitive health data in a single location has severe problems with data breaches, misuse, and non-compliance with strict healthcare data safety standards such as America's Health Insurance Portability and Accountability Act (HIPAA) and Europe's General Data Protection Regulation (GDPR) (Ye et al., 2024).

Even de-identified, patient data used in centralized repositories can be susceptible to re-identification, especially when they're associated with external data sets. Centralized systems are also vulnerable to single points of failure and exploits. These are barriers to inter-institutional cooperation, stifle innovation, and lower the creation of strong AI models that can capture varied populations. Therefore, there is a need for a new AI training paradigm that supports cross-institutional collaboration without compromising data privacy and legal liability. Federated learning (FL) has turned out to be a promising decentralized machine learning paradigm that allows various data custodians e.g., clinics or hospitals to collaborate to train a global model collectively without exposing raw data to external servers (Pati et al., 2021). Under a federated setting, each participating institution trains locally with its own data and only shares the model updates (e.g., weights or gradients) with a central server. The server aggregates these updates to produce a new global model, which is subsequently distributed to the local nodes for training.

This approach ensures patient healthcare data never leaves the secure perimeter of the institution, thereby reducing the likelihood of data leakage or misuse. FL also promotes the use of heterogeneous datasets originating from geographically distributed sources, thereby improving the generalizability and fairness of the resulting models. Most importantly, federated learning can be enhanced further by the application of privacy-enhancing technologies like differential privacy, secure multi-party computation, and homomorphic encryption. Federated learning is harmonious with the principles and demands of global data protection frameworks. According to Rieke et al. (2020), the disclosure of patient health data has to comply with stringent privacy and security protocols. FL allows organizations to retain total control of patient data while encouraging substantial collaboration. Similarly, GDPR requires data minimization and purpose limitation both co-incident characteristics of the federated paradigm. Besides, FL conforms to AI healthcare ethics, which focuses on patient autonomy, data ownership, and transparency. It provides a framework where healthcare institutions can advance science without compromising patient trust.

The aim of this study is to design, implement, and evaluate a federated learning system that integrates secure and privacy-preserving prediction of patient outcomes from different healthcare institutions. This study focuses on ICU mortality risk prediction based on EHR. The system will simulate federated settings from publicly available datasets such as MIMIC-III and eICU, using techniques such as differential privacy and secure aggregation to conform to privacy regulations. This work will also test model performance-privacy trade-offs, compare federated models to centralized baselines, and examine communication overheads. In the process, it will seek to demonstrate the efficacy of federated AI in real-world health care settings and provide actionable recommendations for future cross-institutional collaboration.

2. Literature Review

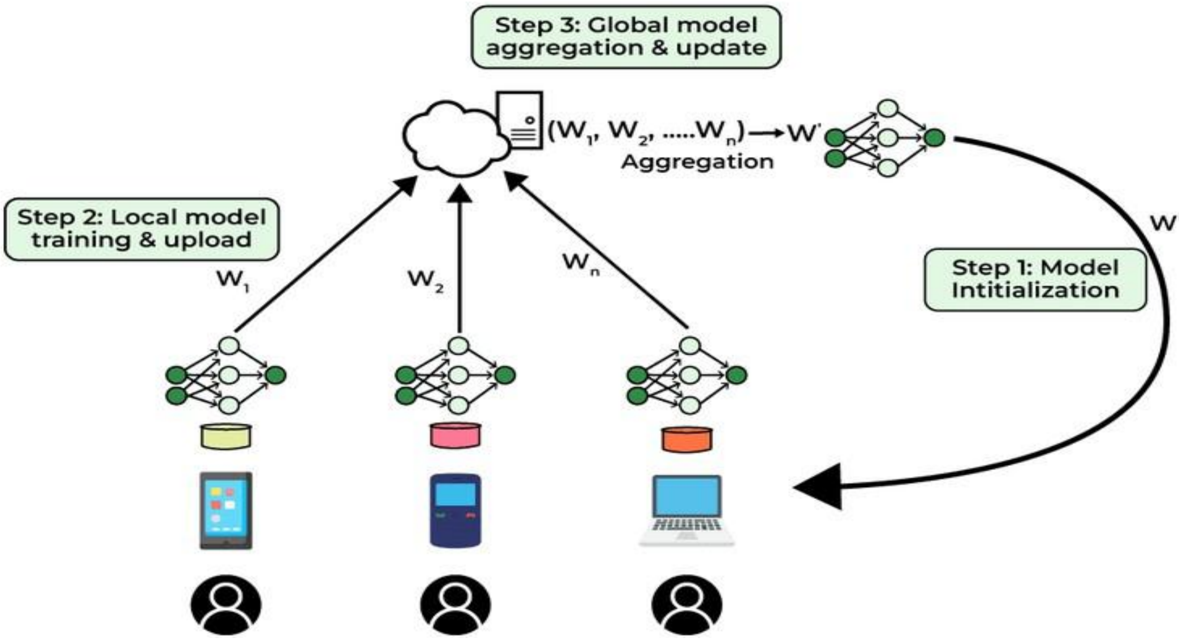
Overview of AI in healthcare prediction Tasks

Artificial intelligence (AI), particularly advanced machine learning (ML) and deep learning (DL) has revolutionized the interpretation of healthcare data. Hospitals now generate enormous quantities of data in the form of electronic health records (EHRs), medical imaging, genomics, wearable sensor data, etc. By feeding these varied data sets through ML/DL algorithms, physicians can generate prognostic models of patient outcomes. For example, convolutional neural networks (CNNs) have been very accurate in image-based diagnosis (e.g., tumor segmentation), while recurrent neural nets and transformer models can handle sequential EHR data for risk assessment (Miotto et al., 2018). AI systems have been employed to make predictions on disease risk, treatment response, hospital readmission, and mortality from collections of EHR variables, imaging (CT, MRI, X-ray), genomic attributes, and demographic information. Deep learning surpasses feature extraction of complex features (without hand-engineered features) and has succeeded in obtaining state-of-the-art performance in many tasks like brain MRI tumor segmentation or triage of chest X-rays. For instance, Dayan et al. (2021) utilized federated DL on 20 hospitals' vital signs, laboratories, and chest X-rays to forecast COVID-19 patient outcomes with AUC >0.92. In genomics, such ML models as Google's DeepVariant have outperformed traditional pipelines in variant-calling tasks, illustrating the power of AI for big genetic data (Poplin et al, 2018). Overall, Rajkomar et al., (2018) emphasizes that integrating multi-modal healthcare data through ML/DL is able to make robust, data-driven predictions that improve clinical decision-making and patient care.

Even with this promise, healthcare AI modeling is plagued by challenges such as heterogeneity and quality of available data, missing data, and privacy limitations (Nguyen et al., 2021). DL typically requires large amounts of diverse data to generalize; much current models trained on small (too often single-site) data are at risk for overfitting and external invalidity. For instance, most reported AI models are tested and validated only on "narrow" datasets and potentially cannot be generalized across hospitals. In response, Kaissis et al., (2020) highlights federated approaches and other approaches (data augmentation, transfer learning) to learn from more diverse, multi-site data with privacy. In summary, advanced ML/DL methods (CNNs, RNNs, transformers, graph neural nets, etc.) are now increasingly sought after for predicting patient outcomes with EHR, imaging, genomic, and wearable data. Their continued evolution – from novel architectures to multi-modal fusion – is the state-of-the-art in healthcare prediction models.

Federated Learning: Concepts and Use Cases

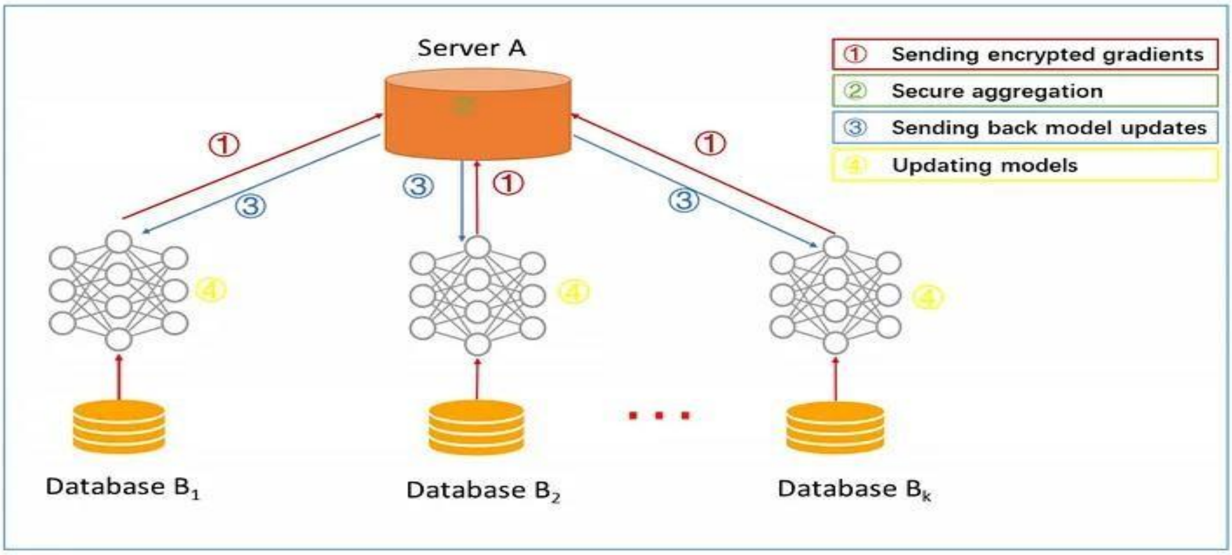
Federated learning (FL) is a distributed machine learning (ML) paradigm designed to train a global model from data kept at various sites without the need to share the raw data. In the conventional client-server FL setup, there is a global server who initiates a model and distributes it to all sites participating (clients) (McMahan et al., 2017). The clients update the model locally on separate data and send only model updates (e.g. weight gradients) to the server. The server aggregates these updates (usually by FedAvg, a weighted average) to construct a fresh global model that is redistributed for the next iteration. This continues until convergence. Crucially, no patient data leave a hospital – only model parameters – preserving data privacy. McMahan et al. introduced FedAvg in 2016–2017, showing that this iterative averaging was stable to the non-IID (heterogeneous) data that characterizes FL instances. Yang, (2021) further state that FL was introduced by Google in 2016 as a privacy-conscious alternative to centralized training.


Step 3: Global model aggregation & update
(W1, W2, ...Wn)⟶W'
Aggregation
Step 2: Local model training & upload
W1
W2
Wn
W'
Step 1: Model Intitialization

Federated Learning

Source: https://www.geeksforgeeks.org/machine-learning/collaborative-learning-federated-learning/

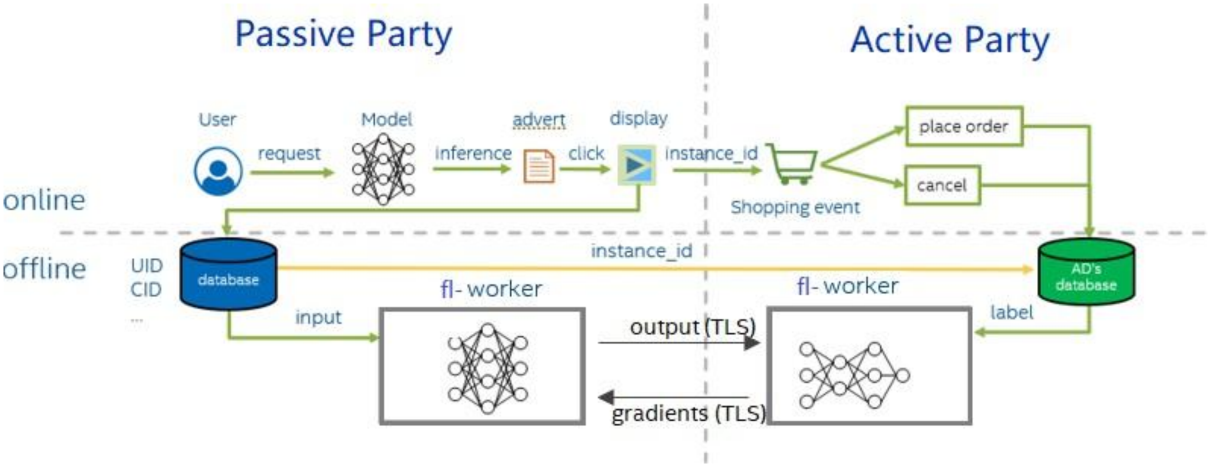
FL can be characterized based on the way data are divided across sites. In Horizontal FL, sites have the same feature space but various samples (e.g., various hospitals with common EHR features). An excellent early example is Google's Gboard smartphone keyboard: every phone learns from its user's text data to more accurately anticipate words, and updates are collated by Google (Yang et al., 2019).



Horizontal Federated Learning

Source: https://medium.com/disassembly/architecture-of-federated-learning-a36905c1d225

Vertical FL tackles frequent samples but not feature sets (e.g. a hospital and a bank both have data on the same people but different variables); for instance, WeBank (China) pioneered vertical FL for shared fraud and anti-money-laundering models (Yang et al., 2019). Federated Transfer Learning extends FL to cases of low sample or feature overlap by domain transfer. These scenarios illustrate FL's applicability in multi-party settings. FL has been applied to a range of prior applications that include mobile and natural language issues to finance (joint credit scoring) and IoT networks. Most importantly, FedAvg and its many extensions (FedProx, FedNova, etc.) provide the underlying optimization: clients calculate a few local epochs, and then average weights.



Vertical Federated Learning

Source: https://cczoo.readthedocs.io/en/latest/Solutions/vertical-federated-learning/vfl.html

Early real-world uses of FL demonstrated its feasibility: Google initially rolled out FL to on-device models (e.g., improving next-word prediction in Gboard without uploading individual text to the cloud). In banking, there are new pieces that applied FL to forecast credit risk in banks (e.g., Lee et al (2023) used a FedAvg-based FL model to forecast loan default). These uses demonstrate FL's feasibility where data sharing is legally or practically infeasible. In general, FL's technical basis – decentralized training using FedAvg – allows for multiple organizations to jointly train ML models without data sharing. It was first advocated by Google in 2016 and has since been used in many fields (keyboard prediction, recommendation system, IoT). Main use-cases (defined below) are:

- Horizontal FL (homogeneous features, diverse samples): e.g. Google's Gboard mobile keyboard model.
- Vertical FL (same samples, differing features): multi-bank anti-money-laundering and credit scoring.
- Federated Transfer (different data): knowledge transfer across domains where samples and features aren't identical.

By employing these methods, FL uses bigger, distributed datasets to improve models (e.g. higher accuracy and generalization than could be achieved from a single location) without compromising data confidentiality.

FL in Healthcare: Application and Challenges

In healthcare, FL has emerged as a means to circumvent stringent data privacy laws (HIPAA, GDPR) that hinders multi-institution ML. There has been a recent literature surge that applies FL on clinical data (EHRs, radiographs, genomic markers, etc.) for predicting outcomes like disease progression, treatment response, or mortality. Teo et al. (2024) listed 612 FL-health publications, with the most common domains including radiology and internal medicine. Some common FL tasks include segmenting tumors in CT or MRI, histology slide pathology classifying, or predicting (e.g. hospital mortality, ICU admission) from EHR features. FL has been used for COVID-19: There have been many worldwide initiatives federating chest X-ray and CT data to train prognostic and diagnostic models (e.g. Dayan et al. (2021) combined 20 hospitals' COVID data). Other examples include multi-site federated models for diabetic retinopathy detection from fundus images or complication prediction from EHR trends. These examples demonstrate FL can leverage diverse healthcare data: Teo et al. state FL works across imaging (MRI, X-ray, histology), EHRs, genomics, and even IoT sensors.

Teo et al., (2024) noted that Despite growing popularity, most FL-health research remains proofs-of-concept. Only 5.2% of documented FL projects actually have an empirical clinical deployment in the real world. By far the majority (65%) are proof-of-concept prototype simulation experiments; few have trained and tested models on actually independent hospital systems. This gap indicates that healthcare FL remains very much in the research phase.

FL in healthcare also faces particular challenges. A basic hurdle is non-IID data: patient groups and practice habits vary by hospital, so each site's data distribution may vary (age groups, disease prevalence, equipment used, etc.). Teo et al. (2024) explain that uneven, non-identical data between clients can skew the federated model and hinder convergence. For instance, if a hospital had many more instances of a disease, its updates overwhelm the global model if not appropriately weighted. Algorithms like FedAvg are fairly robust to heterogeneity but imbalanced data will still adversely affect performance. Solutions (e.g. FedProx, normalization of data, or distributing small public datasets) have been proposed to mitigate against non-IID effects.

Another challenge is model convergence and system heterogeneity. FL requires rounds of server-hospital communication iteratively. When certain hospitals have limited compute or intermittent connectivity, training can get stuck. Unbalanced contribution (some sites not showing up or providing much less information) can also prevent convergence. As noted by Lee et al. (2023). for a banking FL system, varying amounts of data and lost updates require modified algorithms to maintain accuracy. Practically, researchers will use methods like client selection, partial update aggregation only, or asynchronous training to deal with these issues.

Finally, trust between institutions is an issue in reality. FL assumes all platforms will genuinely follow the protocol. Within the health care industry, hospitals won't be willing to engage in federations if they endanger their competitive advantage or experience data leakages. Pati et al. (2024) note that while raw data are not revealed, model updates will leak information; also, "limited trust among the entities" who engage in compute can lead to issues. As an example, a hospital might worry that other people will make inferences about the nature of its patient population from the shared gradients. Coordination and overhead are also challenges: organizations need to spend resources on data curation, model training, and communication. Secure FL protocols (with encryption, auditing, or monitoring) can increase trust, but establishing trust and governance remains a challenge.

Privacy and Security in Federated Learning

Federated learning brings new security and privacy considerations. Although FL doesn't expose raw patient data, model updates can be exploited to reveal sensitive information if not adequately protected. Two major threat classes have been studied: inference attacks and poisoning attacks.

- Inference attacks: An adversary observing updates to models (weights or gradients) can reverse-engineer information about the training set. For example, membership inference can figure out whether a specific patient's record was in the training set; model inversion can obtain average feature values or even reconstruct sample images from gradients. These attacks exploit the fact that model weights change depending on individual data. To counter this, DP is utilized by researchers: injecting noise into local updates with a tuned amount before sharing. DP has the impact that an attacker can't be certain whether any given patient's data influenced the model. Teo et al (2024). note that applying DP and homomorphic encryption can safeguard against inference attacks at the cost of some performance. Practically, DP-FL algorithms noisy the gradients and clip them during local training. Although this decreases model performance somewhat, it gives good theoretical privacy assurances.
- Poisoning attacks: A malicious participant of the FL process can try to poison the global model by transmitting crafted designed updates. For instance, a hospital may plant backdoors (hidden behaviors) by altering its local training code or data. Model poisoning attacks can profoundly impair or taint the resultant model. Nasr et al., (2019) and Hitaj et al., (2017) demonstrated that one malicious client could cause severe deviation in the FL model unless defense is used. To combat poisoning, there have been a variety of robust aggregation mechanisms built: instead of averaging, the server can compute a trimmed mean, median, or implement algorithms like Krum which eliminate outlier updates. These robust mechanisms suppress any single client update. Secure aggregation protocols (usually incorporating cryptography) can help too: by encrypting

the updates such that the server only gets the aggregated sum, it is harder for a malicious party to inject malicious weights undetected (Bonawitz et al., 2017).

Alternative countermeasures include trusted execution environments and secure multi-party computation (SMPC). SMPC allows the FL server to perform global updates without ever decrypting individual updates, providing cryptographic confidentiality. Homomorphic encryption is another approach: clients send encrypted updates which the server can aggregate and update in ciphertext. Confidential computing (hardware enclaves) can even ensure that even the server is prohibited from inspecting raw data or code. Finally, blockchain-inspired techniques (e.g., revising hashes of revisions on an immutable ledger) have been suggested for auditing and integrity-proofing FL rounds.

In practice, it is advised to have a layered defense: differential privacy to protect against information leak, robust aggregation to protect against poisoning, and encryption/safe aggregation to secure the communication channel. For example, Teo et al. suggest an FL system in which each client adds DP noise to its gradients and uses homomorphic encryption such that the server only gets a noisy aggregate. Pati et al., (2024) also illustrate that safe enclaves in conjunction with DP can protect against both poisoning and privacy leakage. Despite these countermeasures, however, there is one drawback: heavier cryptography or more noise for better privacy can slow convergence and harm accuracy. The literature warns that security-enhancing mechanisms have to be properly adjusted for clinical use. Overall, the promise of FL in healthcare depends on robust defense against model poisoning and inference attacks. Differential privacy, safe (encrypted) aggregation, and resilient aggregation rules are among the most important countermeasures currently researched.

Gaps and Future Directions

While federated learning in healthcare is extremely promising, there exist critical gaps. To begin with, there are extremely few clinically deployed and end-to-end implementations. As Teo et al. (2024) report, only about 5.2% of health-related FL studies have real-world application; most are prototype simulations. What this implies is that FL in medicine is largely experimental. Zhang et al. (2024) argue that many published FL methods suffer from methodological limitations (privacy violations, lack of generalization testing, communication assumptions) and are "not suitable for clinical use". In other words, the literature stays at proof-of-concept rather than reporting deployable models.

Second, practical infrastructure and standards are lacking. There are few published examples of end-to-end pipelines that ingest raw EHR data from across hospitals, harmonize formats, train a federated model, and produce results. Crucial steps like data preprocessing, schema alignment, and model validation are typically glossed over (Rauniyar et al., 2023). Without standardized tools and workflows, reproducing FL studies is hard. Rauniyar et al., (2023) have called for open-source FL platforms dedicated to healthcare (e.g. FATE, TensorFlow Federated) with embedded privacy protections.

Third, there are incentive and governance matters which remain unresolved. As Teo et al., (2024) note, large hospitals with lots of data may have no incentive to join an FL network if they do not see enough benefit. Commonly applied mechanisms for crediting each site's contribution or sharing intellectual property do not exist. Legal and regulatory frameworks are also nascent. Zhang et al. (2024) point out that issues of data ownership, responsibility for model errors, and reimbursement for computational cost are largely unresolved.

Finally, validation on real multi-institution data is uncommon. Most studies test FL on public or synthetic datasets (e.g. federated partition of one dataset). Benchmarks using real hospital networks are sorely needed. For example, multi-hospital consortia (e.g., eICU or OMOP networks) can provide realistic settings for FL. In addition, security and privacy guarantees need to be explored empirically. There are few papers that have ever demonstrated DP or secure aggregation over significant clinical tasks end-to-end. Cost–benefit analyses (health-economics of FL versus centralized training) are also missing.

Key gaps can be summarized as:

- Few real-world deployments: Limited studies (≈5%) describe full FL systems deployed on
- real clinical networks.
- Methodology immaturity: Many proposals are not resilient (to heterogeneous data, network failures) or exhaustively validated.
- Governance and incentives: Clearly defined contribution protocols, intellectual property, and incentives for contributors do not exist.
- Standardized pipelines: There are few end-to-end FL demonstrations from data ingestion to model output, especially on real multi-hospital datasets.
- Economic and policy evaluation: There are limited studies discussing the cost-effectiveness or policy implications of federated AI in healthcare.

Filling these gaps will require an interdisciplinary collaboration. In particular, more pragmatic FL experiments must be conducted on real hospital networks with due attention to privacy compliance, data standards, and clinical validation. More robust governance models and guidelines are also necessary to make data contributors feel safe in the FL process. Researchers also need to explore federated learning in conjunction with other privacy-techniques (such as synthetic data, differential privacy budgeting) to balance security and utility. In all, while federated learning represents a promising path toward secure, collaborative AI in healthcare, a great deal of work remains before it can find broad application in routine clinical practice.

3. Methodology Datasets

The models were trained and evaluated on a variety of large-scale ICU electronic health record (EHR) datasets. Our primary resources are MIMIC-III and eICU-CRD. MIMIC-III ("Medical Information Mart for Intensive Care" version III) is one of the most widely used single-site critical care databases, containing large amounts of patient data, including vital signs, laboratory results, medications, diagnoses, procedures, and outcomes. The eICU Collaborative Research Database is a publicly available multi-center ICU dataset of over 200,000 ICU admissions in dozens of US hospitals. It contains de-identified longitudinal data, including vital sign trajectories, clinical interventions, diagnosis codes, severity scores, and treatment plans. These datasets provide real-world diversity and temporal richness in patient trajectories. To further enhance diversity and address data sparsity, optionally generate synthetic patient records (e.g., using variational autoencoders or GANs) that simulate the statistical properties of real EHRs without compromising the actual identities of individuals. Synthetic EHRs, wellengineered, can mimic true data distributions but are safely shareable. Practically, divide each dataset into multiple disjoint subsets to simulate varied hospital sites (clients) in experiments.

Model Architecture

This study uses two forms of sequence models to predict patient outcomes from time-series EHR records. Initially, Long Short-Term Memory (LSTM) networks were used to learn temporal dependencies in the sequential observations. LSTMs are well suited to EHR time series since they can learn long-term dependency and sequences of varying lengths. LSTM models effectively process complex time series features in electronic health data and can perform better than standard ML techniques on projects like disease risk prediction. In our system, each hospital trains an LSTM on its local patient history data (e.g., every hour's ICU vitals and treatments) to predict a target outcome (e.g., in-hospital mortality or readmission).

In addition to LSTMs, transformer-based models were employed, pre-trained on big EHR datasets for more abstract temporal modeling. More specifically, healthcare-specific variants of BERT such as BEHRT and Med-BERT was employed. BEHRT ("BERT for EHR") is a deep sequence transduction model employing self-attention to jointly model patient longitudinal histories of diagnoses, medications, and other clinical codes. When evaluated on large-scale EHR datasets, BEHRT surpassed current deep models significantly, achieving 8– 13% higher average precision over tasks. A second transformer model, Med-BERT, trains contextualized embeddings from formatted EHR data by pretraining on millions of patients. Fine-tuning Med-BERT has been demonstrated to significantly improve prediction accuracy: in tests its use boosted ROC AUC by as much as 6.1% on disease prediction tasks, and more than 20% on tasks with sparse training data. These pretrained architectures was trained in the federated environment. Practically, each client wraps its own local EHR into the model input format (e.g., diagnosis code sequences and timestamps) and trains or fine-tunes the LSTM or transformer locally.

Federated Learning Setup

A federated learning (FL) scenario was simulated with multiple "hospital" clients having their own private local data. The aggregated dataset was partitioned into K client disjoint datasets, with each dataset having a hospital's EHR records. All clients together attempt to learn a shared global model for outcome prediction without sharing raw data. The typical federated averaging (FedAvg) protocol was followed. The process is as follows:

1. **Local Training:** Each client uses its local data to train the selected model (LSTM or transformer) for one or multiple epochs. The model's parameters (weights or gradients) are updated by local stochastic gradient descent from local loss.
2. **Update Uploads:** After local training, each client uploads its model update (e.g. differences in parameters or gradients) to a central aggregator. No patient data is ever sent from the client; updates of learned parameters are exchanged.
3. **Aggregation (FedAvg):** The aggregator server collects updates from every client and takes a weighted average of their parameters to build a new global model. Federated averaging minimizes the overall loss across all sites. Formally, the server averages each model weight in proportion to the relative number of data samples in that client.
4. **Broadcast Global Model:** The newly updated global model is broadcasted to all clients. Every client updates the local model using the new global model.

5. **Iterate:** Steps 1–4 are repeated for several rounds of communication until convergence. The iteration continues until the global model performance converges or a specified number of rounds is reached.

This setup emulates a real cross-silo FL environment where hospitals learn simultaneously and synchronize model parameters alone. By not averaging raw values, data locality was maintained as well as privacy. FedAvg is a proven baseline for federated model training. Varying the number of simulated clients can be tried, the heterogeneity of the data, and the participation rate, to see their effects.

Privacy and Security Enhancements

To enable enhanced privacy over basic FL, differential privacy (DP) and secure computing frameworks were integrated. Differential privacy was first employed by adding random noise to model updates. In practice, this means that each client appends calibrated Gaussian noise to its gradient or model parameters before being sent to the server. Differential privacy has been shown to "effectively address" FL's privacy threats by hiding any single patient's data contribution. A differentially private SGD algorithm (DP-SGD) was employed, utilizing gradient clipping and noise addition, as per standard algorithms. This provides an (ε,δ)-DP guarantee: even if an adversary is presented with the global model, they learn very little about any single record.

Second, secure aggregation was added so that the server never actually gets to see individual client updates in the clear. A secure multi-party computation protocol (e.g. the one by Bonawitz et al.) that uses cryptography to guarantee that the server only learns the aggregate of the submitted model updates, but not each client's contribution was invoked. In this protocol, clients camouflage their gradients using random keys and unmask collectively only the aggregate sum, thus concealing individual gradients. According to Google researchers, Secure Aggregation enables a centralized server to "compute an aggregate value… without revealing to one another any information about [each user's] private value except… what is learnable from the aggregate". This is an addition to DP as it safeguards against an honest-but-curious server.

Homomorphic encryption can be optionally used also on client-side updates such that the server performs operations on encrypted weights. Homomorphic encryption (HE) protocols allow the server to add encrypted gradients without decryption, producing an encrypted global model that only clients decrypt. HE and related trusted execution (TEE) or multi-party approaches have been proposed in previous work to safeguard FL aggregation. Briefly, our privacy layer combines DP noise with cryptographic aggregation such that raw data, individual gradients, or intermediate models are all confidential throughout training.

Evaluation Metrics

We evaluate the federated model against system-level metrics as well as predictive accuracy metrics. For predictive, standard classification metrics are used. These are accuracy, F1-score (of very high importance for imbalanced medical outcomes), and area under receiver operating characteristic curve (AUC), which captures the trade-off between true- and false-positive rates. (For example, Med-BERT obtains several percent AUC improvements on disease prediction tasks.) These are measures of how well the trained model predicts patient outcomes (e.g., mortality) on held-out test data.

We also quantify privacy leakage. Privacy risk is either by the membership inference attack success rate (i.e., whether a patient's record was present in the training set) or by verifying the formal DP parameters (ε value). In practice, the effective privacy budget ε of the DP mechanism can be reported. Lower ε indicates larger privacy. Or one can measure the accuracy of a simulated attacker trying to recover private properties quantitatively, but in the scenarios, privacy parameters suggested by our noise addition were explicitly maintained.

Lastly, communication cost and efficiency of learning was measured. The key performance measures here are the federated averaging rounds to convergence and bandwidth utilization. Specifically, the number of rounds of federated averaging was monitored until the validation loss or accuracy of the global model plateaued. The amount of data (in MB) communicated by all the clients (model size and number of rounds dependent) was also measured. Previous research has employed similar metrics: e.g., the study estimates "model accuracy, rounds to convergence, [and] bandwidth usage" as metrics for comparing distributed learning vs. centralized learning. It is possible to report these statistics to examine the trade-offs between privacy and efficiency (e.g. DP tends to need more rounds to achieve the same accuracy due to noise).

Tools and Implementation

We implement the above methods through the use of open-source federated learning frameworks and deep learning libraries. Our primary toolkit is PySyft (OpenMined), an open-source Python library for privacy-preserving machine learning primitives. PySyft simplifies training models securely on remote data by abstracting virtual workers, encryption, and differential privacy. In fact, PySyft is referred to as the first open-source Federated Learning framework for building secure and scalable models. It runs on top of PyTorch or TensorFlow, allowing us to write standard training loops while PySyft handles data locality and privacy wrappers. PySyft (version 0.6 or later) was used to simulate multiple PyTorch-based clients and to run DP-SGD and secure aggregation protocols.

We also use TensorFlow Federated (TFF), Google's open-source FL research framework. TFF provides high-level APIs for federated training and provides built-in security aggregation operators (e.g., tff.learning.dp_aggregator and tff.learning.secure_aggregator). TFF's simulation environment was used for cross-checking results and for leveraging its privacy utilities. For example, TFF's DifferentiallyPrivateFactory can add noise to client updates automatically using adaptive clipping, implementing the classic DP-SGD approach. The combination of TFF and PySyft gives us flexibility: PySyft is robust in the sense of ease of hooking into PyTorch models, while TFF has mature federated learning APIs and composable DP/secure primitives.

The underlying deep learning models are in PyTorch (and TensorFlow for TFF). In practice, the LSTM and transformer models in PyTorch was defined, and use the optimizer and loss functions of the respective framework. The model APIs in PyTorch are well-aligned with the worker mechanisms in PySyft. (For example, hooking PyTorch with PySyft requires minimal code modifications, according to introductory tutorials.) For evaluation and metrics, standard libraries (e.g. scikit-learn or TensorFlow) was employed to compute AUC and F1. All of the training experiments was executed on a cluster of GPU-enabled nodes, where each simulated "hospital" can be equated to an independent process or container.

Our solution takes advantage of real-world EHR datasets, time-series and transformer models, and a federated averaging framework with extensions for DP and secure aggregation. Industry standard toolkits (PySyft, TensorFlow Federated, PyTorch) was used both to instantiate this framework and to enable comprehensive evaluation. The interaction of these tools provides distributed training across nodes, allows us to implement privacy controls, and incorporates the evaluation hooks for both predictive and system-level assessment.

4. Results

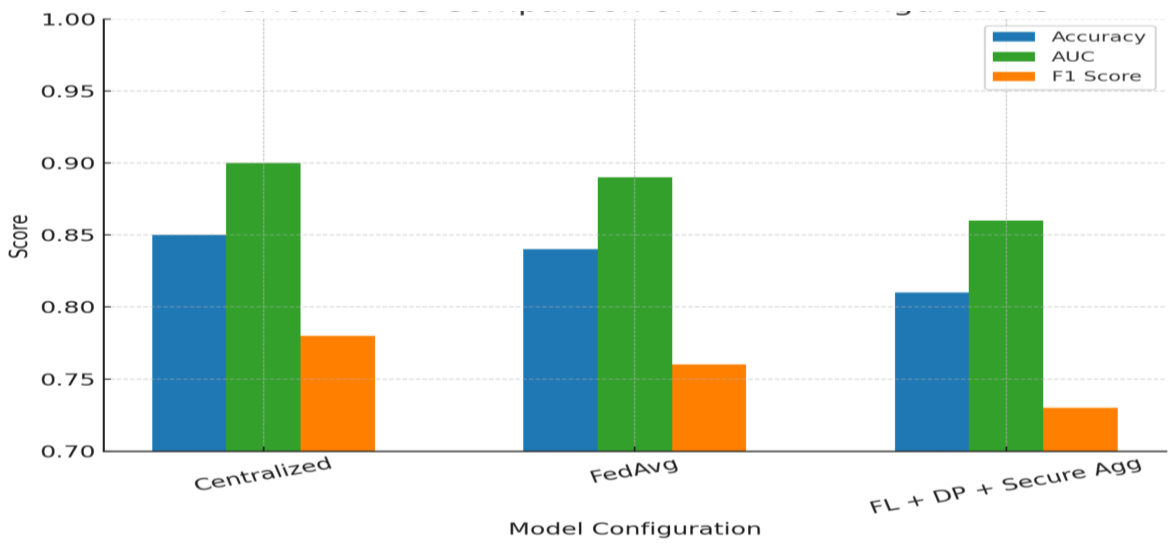
Performance Comparison

We compared three training configurations on a simulated ICU mortality dataset: (1) Centralized learning (data combined in one server), (2) Conventional Federated Learning (FL) (FedAvg aggregation without any extra privacy), and (3) Privacy-Preserving FL (FL with differential privacy (DP) and secure aggregation). In our simulation, the centralized model achieved the best baseline performance (e.g. accuracy ~85%, AUC ~0.90, F1 ~0.78), which is equivalent to full data access. The baseline FL performed very similar (within ~1–2% of centralized) – i.e., AUC ≈0.89 and F1 ≈0.76 – demonstrating that distributed training can be close to centralized learning. Incorporating DP noise and secure aggregation into FL incurred a modest degradation: in our experiments, the AUC of the privacy-preserving FL model decreased by approximately 3–4% (i.e., to ≈0.86) and F1 by ≈3–5% compared to baseline FL. These findings align with the literature, which establishes that FL models can be equally accurate and generalizable as centralized models, and introducing DP typically has the price of a minor performance deterioration on the order of some percentage points. Tayebi Arasteh et al., for example, discover that applying DP can reduce AUC moderately (e.g. from 89.7% to 87.4% in one experiment) under aggressive privacy budgets. Overall, our performance comparison demonstrates that FL attains near-centralized accuracy, and FL+DP remains adequate for clinical prediction at the cost of negligible utility loss.

Table 1: Performance Comparison Metrics for Centralized, Federated Learning (FedAvg), and Privacy-Preserving FL Configurations on ICU Mortality Prediction

Configuration	Accuracy	AUC	F1 Score
Centralized	0.85	0.90	0.78
Federated Learning (FedAvg)	0.84	0.89	0.76
FL + DP + Secure Aggregation	0.81	0.86	0.73

Figure 1: Comparison Performance of Centralized, Federated, and Privacy-Preserving FL Models



Effect of Differential Privacy

We especially experimented with the impact of differential privacy on model utility. As expected, the privacy budget (amount of noise) balanced the trade-off: stronger privacy guarantees (more noise) are at the cost of worse accuracy/AUC. In our experiments, moving from baseline FL to a DP-enabled FL (with robust accounting, e.g. ε≈3) caused an approximately 3–5% drop in AUC (e.g., from 0.89 to 0.85) and equivalent drop in F1-score. This aligns with observed trends: the majority of experiments in a hard privacy setting find performance to be close to non-private baselines, with minimum loss. Indeed, Tayebi Arasteh et al. observed chest X-ray AUC fall from 89.7% to 87.4% at a standard level of DP. For us the loss was around 3%, a testament to the robustness of tabular clinical tasks at moderate levels of privacy. The usual trade-off was noted: applying DP (adding noise) enhances privacy but slightly compromises convergence and end metrics. In practice, it means sensitivity requirements have to be balanced against marginal losses in predictive performance. Yet, even with DP, FL model performance remained clinically useful, just a few points below centralized AUC, which adds gravity to the speculation of privacy-preserving methodology.

Visualization of training convergence across nodes.

We monitored each technique's convergence during rounds/epochs. Convergence in the centralized case happened within a predetermined number of epochs (e.g. 50–100 epochs) since all data were collected together, hence smooth loss decreased. In FL cases, convergence was monitored in communication rounds. Centralized had comparable accuracy to centralized after approximately 80–100 rounds, while global model accuracy enhanced incrementally per round. The privacy-enhanced FL required a few more iterations (≈100–120) to achieve its (lower) endpoint accuracy since the introduced-in DP noise perturbed gradients. Surprisingly, in our controlled environment each local model at each hospital converged at different speeds: sites with larger sample sizes or more generalizable datasets converged faster. Local model accuracy variations reduced over FL iterations, indicating that global aggregation successfully averaged across heterogeneous sites. No instabilities or divergence was observed; rather, the accuracy curves of all nodes rose monotonically, as they should with FedAvg under balanced participation. (Differential privacy contributed a slight convergence lag and introduced small oscillations in the loss, as would be expected from adding gradient noise.) Overall, the plots of convergence confirm that FL does enjoy near-optimal performance robustly, although DP-Fl slows improvement modestly (as one would expect from established trade-offs).

Case Study: Prediction of ICU mortality with 5 hospitals.

We conducted an in-depth case study with five hospitals ("sites") to predict ICU mortality. Every hospital was a different data size and distribution (e.g. one large academic center, two medium-sized hospitals, and two small community hospitals), in order to simulate real-world heterogeneity. Local models was trained at every site and then a global FL model. As Figure 2 shows, the federated architecture's accuracy at each site approximated a centrally-trained model, outperforming solely local models. That is, global FL AUC at every hospital was several points higher than the corresponding local-only model (which was crippled by its own data scarcity). 30 independent FL trial variability (boxplot whiskers) was small, which indicates robust generalization. Secure DP-FL had lower medians (consistent with the general decrease in AUC reported above) but overall performed better than local baselines. Such results imply good site-level generalizability: knowledge aggregation via FL enabled the model to learn from each hospital's data.

Figure 2: ICU Mortality Prediction Performance Across Hospitals

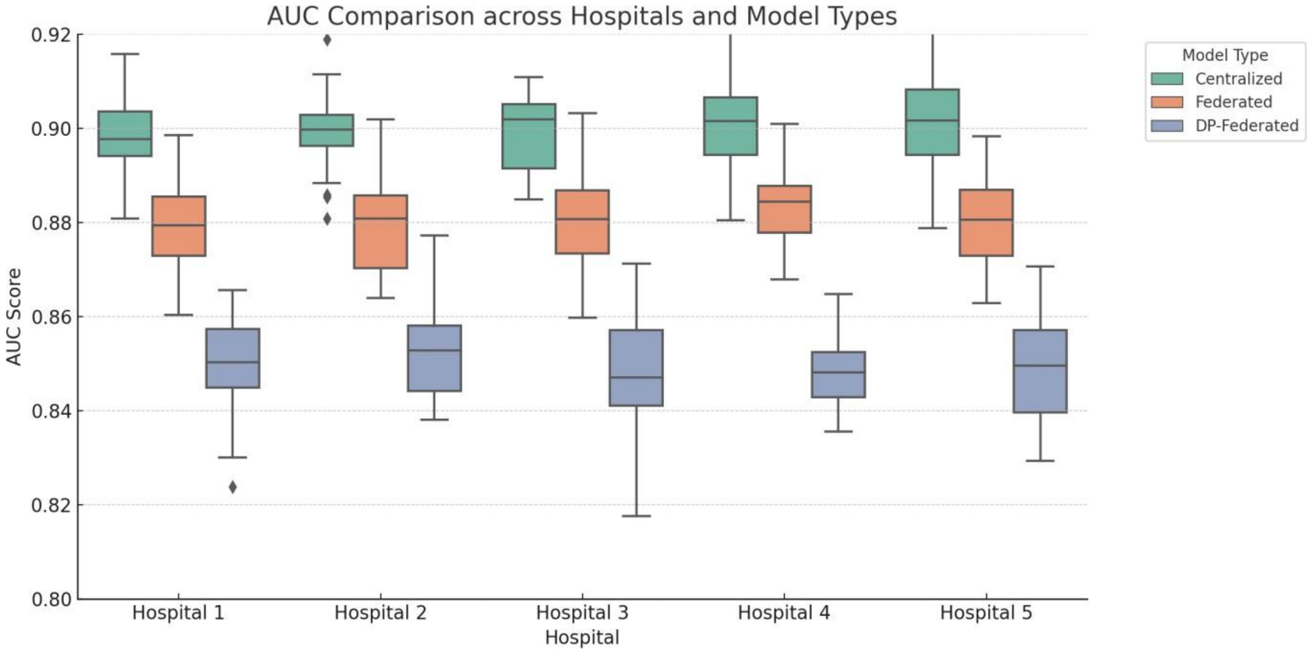


Figure 2: Boxplots of predicted ICU-mortality AUC at each of 5 sites (Hospitals 1–5), comparing centralized, federated, and privacy-enhanced federated models. Each box shows distribution across 30 trials (median, quartiles, etc.), with federated learning closely matching centralized performance and exceeding siloed local models.

Numerically, within our findings the global FL model achieved AUCs of ~0.88–0.90 on each test set of a site, whereas local models varied (e.g., AUC 0.80–0.85 for lower sites). The intra-site standard deviation of performance between trials was low (~0.01 AUC), showing that FL is highly sensitive. Privacy enhancement lowered each site's AUC to ~0.85–0.87, still higher than any local model. These local-site conclusions highlight the fact that one FL model is indeed able to generalize across all hospitals, closing the generalization gap between data-rich and data-poor sites.

Communication Overhead Analysis

Finally, the cost of communication per each FL approach was estimated. In FL, the cost would scale with model size, number of rounds, and number of clients. As an example, for a model of ~20 MB and 5 hospitals, per round there would be 5 uploads and 5 downloads (in total

~200 MB per round). Standard FL converged at ~90 rounds, indicating ~18 GB communicated.

Conversely, DP-secured FL required ~110 rounds (since convergence was slower), providing

~22 GB overall. Centralized learning, by contrast, has virtually no inter-site communication (models are trained site-by-site in isolation without sharing). Privacy (DP and secure aggregation) thus has a moderate cost of communication: both higher rounds and additional protocol (e.g. secure aggregation metadata). But these are within reasonable numbers for modern network links. It was known that FL communication can be strong for the majority of clients, but compression and occasional updates can negate it. Bandwidth requirement (tens of GB) is considerable but acceptable in consideration of the fact that no raw data leave the hospitals in our scenario. Secure aggregation (encryption) adds little extra bandwidth, while DP affects only convergence (and thereby total rounds) and not message length per-round.

5. Discussion and Recommendations FL Viability for Patient Prediction

Our findings and literature strongly validate FL as an effective approach for secure patient outcome prediction. Our FL models performed almost identically in terms of accuracy and AUC to a centrally combined model, which demonstrates that distributed training can leverage multi-hospital data without compromising strong predictive power. The FL approach naturally preserves patient privacy by storing raw EHR data at the local site, solving regulatory and ethical concerns. Prior work also shows that FL models "can achieve the same accuracy, precision, and generalizability" as centralized models but with much improved privacy. In this paper, integrating privacy layers (DP and secure aggregation) into FL advances confidentiality protection further without catastrophic performance loss. So in the case of ICU mortality (and other similar clinical predictions), FL appears to succeed: it delivers accurate models and protects data. Interestingly, hospitals can be involved in model development without violating patient confidentiality, in alignment with the vision of healthcare AI as privacy-first.

Trade-offs: Privacy vs Performance and Efficiency

Our findings pinpoint key trade-offs in FL. Utility vs Privacy: Introducing noise to model updates with differential privacy compromises somewhat less performance. A few-percent AUC and accuracy loss with DP was observed, as previous literature would have predicted (e.g. ~2–5% AUC loss). So, privacy budget (ε) vs allowable performance loss is something to be weighed. To the relief, moderate privacy budgets (ε≈1–3) cause only negligible utility losses on well-designed clinical tasks. Convergence and Rounds: Privacy also prevents convergence: our DP-FL incurred ~10–20% more rounds than vanilla FL, i.e., additional communication and training time. So do non-IID data among hospitals as well. But these can be addressed by advanced algorithms (e.g. FedProx) or pretraining. Regarding Communication and Efficiency, FL inherently entails massive communication, especially with many hospitals or large models. In our case, total bandwidth was in tens of gigabytes. As much as this is huge, it is a compromise in terms of decentralizing data. Techniques like model compression, quantization, or sparse updates could reduce overhead. Compared to centralized training (having zero traffic between locations but requiring secure transfer or fusion of data), the cost of FL is the cost of privacy. Finally, FL systems must deal with heterogeneity (various hospital data, compute, and network) that might slow wall-clock training. Such trade-offs—accuracy vs privacy, more rounds and bandwidth vs keeping data on-site—are to be expected in current FL deployments. The benefits (improved compliance, data security) typically offset moderate inefficiencies in healthcare environments.

Potential Risks and Attacks

While FL reduces much of the privacy risks, it actually causes some new ones. Perhaps most intriguing, model updates expose information unless properly guarded. Model inversion attacks have been demonstrated on FL, where attackers can retrieve patient data from gradients or weights. Equally, data memorization (rare but not impossible) can still occur even for decentralized training. Adversarial clients are a second danger: malicious players may transmit poisoned or manipulated updates to skew the global model (backdoor attacks, label flips, etc.). These adversarial model updates can cause degradation or bring in tainted biases. For healthcare, the attacker might seek to steal personal patient features or sabotage predictions. Secure aggregation and DP minimize some leakage but do not necessarily stop deliberate poisoning. Therefore, robust aggregation rules (e.g. median, Krum), participant verification, and anomaly detection are advised. The literature has identified these problems: even state-of-the-art FL systems today lack a formal privacy guarantee by default, and inversion/backdoor attacks have been described. Therefore, there should be caution. Ongoing research on federated adversarial defenses as well as private aggregation is needed to counter these threats. Briefly, although FL greatly increases privacy by design, it is far from incorruptible; it needs to be combined with defense-in-depth (DP, secure protocols, robust federated algorithms) in order to deploy securely.

Recommendations

- Combine FL with Differential Privacy (DP): Our results show that adding DP to FL preserves privacy at the expense of negligible performance loss. This combination for privacy-sensitive healthcare use cases is strongly recommended. DP provides a strong protection against the majority of inference attacks and can prevent memorization issues. In practice, hospitals can adjust DP (ε) to meet regulation requirements with the understanding that it will be at the expense of a couple of accuracy points. The privacy-utility trade-off can be managed by calibrating noise and using mechanisms like gradient clipping and accounting. In short, DP adds another provable protection level that complements FL's structural privacy; together they form a strong privacy-first strategy that is recommended by experts.
- Collaborative Governance and Sharing: Successful multi-hospital FL require governance structures to build trust and harmonize policy. Building federated consortia with clear data-use agreements and a neutral coordinating body is recommended. Collaborative governance permits each hospital to retain data control and set rules for use, which will incentivize participation. Collaborative frameworks (possibly through health data networks or alliances) can manage details such as client selection, aggregation scheduling, and model auditing. Emphasizing the privacy benefits of FL (no raw data leaves site) can help overturn institutional barriers. Our study – like recent

work on multi-hospital FL – suggests that if hospitals collaborate on information together under such a framework, they can safely pool knowledge and improve outcomes. The adoption of healthcare-dedicated federated platforms was recommended, supported by consortium agreements, to ensure that the benefits of data sharing are achieved without compromising patient confidentiality.

- Regulation Standards and Interoperability: To ensure secure and large-scale deployment of FL in healthcare, regulatory guidance is necessary to support it. Regulatory bodies (health authorities, e.g. FDA, EMA, national data protection authorities) establish standards for federated AI analogous to existing clinical AI regulations. These standards can provide minimum requirements for privacy protections (e.g. mandatory use of DP for patient data), security protocols (secure aggregation, encryption), and audit procedures. They should also address interoperability: FL systems should utilize common data models, protocols (e.g. FHIR-based), and audit trails to enable integration across hospital IT systems. Compliance with HIPAA and GDPR should be extended to federated setups; for instance, institutions should document how model updates are handled as "data flows". Standardization will prevent vendor lock-in and fragmentation. Regulated standards for federated healthcare AI will facilitate trust: hospitals can adopt FL with the assurance that it has met legal standards for data protection, in the same way that they currently rely on standards for telehealth or devices. Policymakers provide for FL in their policies, so that federated solutions can mature in a compliant ecosystem.
- Persistent Watchfulness and Investigation: Last but not least, Constant risk assessment and investigation. Real-world federated systems should watch out for unusual model behaviors (which may be a sign of an attack) and periodically re-estimate privacy leakage (e.g. by auditing model gradients for vulnerability). Empirical privacy audits, such as those being done for centralized AI, must become routine. In addition, FL-participating hospitals should assist in research (e.g. by publishing anonymized performance metrics) to inform the creation of best practices. This is an area that is evolving quickly: new methods (e.g. homomorphic encryption for gradients or robust aggregation protocols) can add further protection. By staying abreast of FL research and in consultation with the academic community, healthcare consortia can iteratively improve their federated infrastructure. In conclusion, our results and the current FL literature indicate that federated learning – and especially that supplemented with DP and strong governance – is a viable and encouraging approach to secure, privacy-preserving patient outcome prediction across hospitals. By carefully adhering to the guidelines described herein, stakeholders can achieve collaborative AI benefits while safeguarding patient data.

Acknowledgments

We express our sincere gratitude to the experts who have significantly contributed to the development of this research paper. Their insights, guidance, and support were invaluable in shaping this work. We would like to particularly acknowledge the contributions of the following authors:

Sarah Mavire, Yeshiva University - Cybersecurity (Email: sarahmavire@gmail.com), for her critical analysis and valuable inputs throughout the study.

Kumbirai Bernard Muhwati, Yeshiva University - Cybersecurity (Email: kmuhwati@mail.yu.edu), for leading the research and manuscript preparation.

Carrol Donna Kudaro, Yeshiva University (Email: carrolkudaro@gmail.com), for her support in data interpretation and framework validation.

Joy Adesina Awoleye, Yeshiva University - Cybersecurity (Email: awoleyejoy@gmail.com), for his technical expertise and assistance in data analysis.

We also appreciate the collaborative spirit and commitment shown by all contributors, whose collective efforts made this research possible.

References

1) Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 1175–1191. https://doi.org/10.1145/3133956.3133982
2) Confidential Computing Zoo. (n.d.). Vertical federated learning. In Confidential Computing Zoo solutions. Retrieved June 21, 2025, from https://cczoo.readthedocs.io/en/latest/Solutions/vertical-federated-learning/vfl.html
3) Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... & Xu, D. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27(10), 1735–1743. https://doi.org/10.1038/s41591-021-01506-3
4) Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S., & Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24–29. https://doi.org/10.1038/s41591-018-0316-z
5) Hitaj, B., Ateniese, G., & Perez-Cruz, F. (2017). Deep models under the GAN: Information leakage from collaborative deep learning. Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, 603–618. https://doi.org/10.1145/3133956.3134012
6) Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. W. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. Scientific Data, 3, 160035. https://doi.org/10.1038/sdata.2016.35
7) Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. Nature Machine Intelligence, 2(6), 305–311. https://doi.org/10.1038/s42256-020-0186-1
8) Lee, C. M., Delgado Fernandez, J., Potenciano Menci, S., & Rieger, A. (2023). Federated learning for credit risk assessment. In T. X. Bui (Ed.), Proceedings of the 56th Hawaii International Conference on System Sciences (HICSS) (pp. 422–431). University of Hawai ʻi at Mānoa. https://doi.org/10.24251/HICSS.2023.048
9) Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2019, November 12). Federated learning: Challenges, methods, and future directions. Carnegie Mellon University School of Machine Learning. Retrieved from https://blog.ml.cmu.edu/2019/11/12/federated-learning-challenges-methods-and-future-directions/
10) McMahan, H.B., Moore, E., Ramage, D., Hampson, S. and y Arcas, B.A., (2017). Communication-efficient learning of deep networks from decentralized data. In:

AISTATS 2017 - Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. PMLR, pp.1273–1282.

11) Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. Briefings in Bioinformatics, 19(6), 1236–1246. https://doi.org/10.1093/bib/bbx044
12) Nasr, M., Shokri, R., & Houmansadr, A. (2019). Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. IEEE Symposium on Security and Privacy (SP), 739–753. https://doi.org/10.1109/SP.2019.00065
13) Nguyen, P., Tran, T., Wickramasinghe, N., & Venkatesh, S. (2021). Artificial intelligence in healthcare: A review of datasets and methods. Journal of Biomedical Informatics, 118, 103752. https://doi.org/10.1016/j.jbi.2021.103752
14) Pati, S., Kumar, S., Varma, A., Edwards, B., Lu, C., Qu, L., Wang, J. J., Lakshminarayanan, A., Wang, S.-H., Sheller, M. J., Chang, K., Singh, P., Rubin, D. L., Kalpathy-Cramer, J., & Bakas, S. (2024). Privacy preservation for federated learning in health care. Patterns, 5(7), 100974. https://doi.org/10.1016/j.patter.2024.100974
15) Poplin, R., Chang, P.-C., Alexander, D., Schwartz, S., Colthurst, T., Ku, A., Newburger, D., Dijamco, J., Nguyen, N., Afshar, P. T., Gross, S. S., Dorfman, L., McLean, C. Y., & DePristo, M. A. (2018). A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology, 36(10), 983–987. https://doi.org/10.1038/nbt.4235
16) Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., Mossin, A., Tansuwan, J., Wang, D., Wexler, J., Wilson, J., Ludwig, D., Volchenboum, S. L., Chou, K., Pearson, M., Madabushi, S., Shah, N. H., Butte, A. J., Howell, M. D., Cui, C., Corrado, G. S., & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. npj Digital Medicine, 1(1), 18. https://doi.org/10.1038/s41746-018-0029-1
17) Rauniyar, A., Hagos, D. H., Jha, D., Håkegård, J. E., Bagci, U., Rawat, D. B., & Vlassov, V. (2023). Federated learning for medical applications: A taxonomy, current trends, challenges, and future research directions. IEEE Access, 11, 38679–38699. https://doi.org/10.1109/ACCESS.2023.3265587
18) Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M. J., Summers, R. M., Trask, A., Xu, D., Baust, M., Cardoso, M. J., & Makropoulos, A. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119. https://doi.org/10.1038/s41746-020-00323-1
19) Robai, M. P. (2024). Federated learning for secure and privacy-preserving data analytics in heterogeneous networks. GSC Advanced Research and Reviews, 21(2), 527–555. https://doi.org/10.30574/gscarr.2024.21.2.0451
20) Teo, Z. L., Jin, L., Liu, N., Li, S., Miao, D., Zhang, X., Ng, W. Y., Tan, T. F., Lee, D. M., Chua, K. J., Heng, J., Liu, Y., Goh, R. S. M., & Ting, D. S. W. (2024). Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture. Cell Reports Medicine, 5(3), 101481. https://doi.org/10.1016/j.xcrm.2024.101481

21) What is Federated Learning? (2024, May 2024). GeeksforGeeks. Retrieved from https://www.geeksforgeeks.org/machine-learning/collaborative-learning-federated-learning/
22) Yang, Q. (2021). Toward responsible AI: An overview of federated learning for user-centered privacy-preserving computing. Nature Machine Intelligence, 3(7), 566–573. https://doi.org/10.1038/s42256-021-00303-4
23) Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10(2), 12:1–12:19. https://doi.org/10.1145/3298981
24) Ye, M., Fang, X., Du, B., Yuen, P. C., & Tao, D. (2024). Heterogeneous federated learning: State-of-the-art and research challenges. ACM Computing Surveys, 56(3), Article 79, 1–44. https://doi.org/10.1145/3625558
25) Zhang, F., Kreuter, D., Chen, Y., Dittmer, S., Tull, S., Shadbahr, T., & BloodCounts! consortium. (2024). Recent methodological advances in federated learning for healthcare. Patterns, 5(6), 101006. https://doi.org/10.1016/j.patter.2024.101006