International Journal of Engineering Research & Management Technology
(Peer-Reviewed, Open Access, Fully Refereed International Journal) ISSN: 2348-4039
Volume 13, Issue-4 July-August- 2026 Impact Factor: 7.09
Email: editor@ijermt.org www.ijermt.org

FEDERATED LEARNING FOR PRIVACY-PRESERVING AI IN HEALTHCARE AND FINANCE

Parul Berwal

Assistant Professor

Shri Vishwakarma Skill University Dudhola

Deepak Singh

Assistant Professor

Shri Vishwakarma Skill University Dudhola

Abstract

The expansion of artificial intelligence (AI) within healthcare and financial sectors has been limited by a core conflict: predictive algorithms excel when developed using extensive, varied datasets, but the essential patient information and transaction records are classified as highly sensitive personal data, governed by stringent regulations like the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR). Federated Learning (FL) has surfaced as a viable alternative, enabling several hospitals, clinics, or financial entities to jointly develop a common model while maintaining the decentralization of raw data. This document offers a narrative literature review on federated learning within the domains of healthcare and finance, succeeded by an innovative simulation-based experimental investigation that applies the Federated Averaging (FedAvg) algorithm across non-homogeneously distributed client shards and evaluates its performance against centralized and isolated-silo benchmarks. Employing a diagnostic breast-cancer dataset as a healthcare proxy and a synthetic, class-imbalanced transaction dataset as a financial proxy, the research reveals that federated training achieves 96–98% of centralized accuracy in healthcare without necessitating raw data transfer, while in finance, federated learning provides a superior precision-recall equilibrium compared to standalone local models, albeit with a quantifiable expense when differential-privacy noise is applied. The study determines that federated learning is a feasible and regulation-compliant approach for inter-institutional AI; nonetheless, its implementation hinges on meticulously balancing the privacy-utility trade-off, statistical diversity, and communication costs that this research measures.

Keywords

Federated Learning; Privacy-Preserving AI; Healthcare Data Privacy; Financial Fraud Detection; Differential Privacy; Distributed Machine Learning; Data Security; Regulatory Compliance (HIPAA/GDPR)

1. Introduction

1.1 Background and Motivation

Machine learning systems' efficacy is contingent upon the quality of the data utilized for training, and in sectors like healthcare and finance, the most valuable data is often the most confidential. The electronic health records (EHR) of an individual hospital seldom encompass a sufficient variety of cases to develop a model that is applicable across different populations, imaging devices, or patterns of comorbidity; likewise, a single bank only observes a small segment of the transaction network where financial fraud transpires, as fraudulent individuals intentionally distribute their activities among various institutions to evade detection. In both fields, the statistically optimal approach — aggregating data from numerous organizations into a singular training set — is exactly what data protection legislation and institutional regulations are designed to prohibit.
Federated Learning, initially articulated by McMahan et al. (2017) using the Federated Averaging (FedAvg) algorithm, redefines the issue: rather than transferring data to a centralized model, the model is transported to the data. Every participating institution develops a localized version of a communal model utilizing its own dataset and transmits solely the revised model parameters (weights or gradients) back to a central server, which consolidates them—usually thru a weighted average—into an enhanced global model. Unprocessed records remain inside institutional boundaries. Kairouz et al. (2021) and Li, Sahu, Talwalkar, and Smith (2020) later expanded this concept into a comprehensive research agenda encompassing communication efficiency, statistical heterogeneity (non-IID data), system heterogeneity, and privacy threat models, thereby defining FL as a unique sub-discipline of distributed machine learning rather than merely a singular algorithm.

Healthcare and banking are commonly perceived as the two sectors where FL's value proposition is most robust, mostly due to their simultaneous demand for data and stringent regulations. Rieke et al. (2020) contend that federated learning (FL) can harness the "untapped potential" of isolated medical data for applications like tumor segmentation, disease forecasting, and outcome prediction, whereas an increasing volume of finance-related studies (Alhasawi et al., 2025; Emmanuel, 2025) demonstrates that FL allows banks to collaboratively enhance fraud detection models while safeguarding customer privacy and avoiding competitive data-sharing issues. Both bodies of literature converge on a shared warning: federated learning (FL) does not inherently eradicate privacy risks; it simply obviates the necessity for raw data sharing, and residual threats like gradient inversion, membership inference, and model-update leakage require proactive mitigation, often thru supplementary methods such as differential privacy (DP), secure multi-party computation (SMPC), or homomorphic encryption (HE).

1.2 Problem Statement

Notwithstanding the extensive and rapidly expanding corpus of conceptual and survey material regarding federated learning in the domains of healthcare and finance, three deficiencies inspire this paper. Primarily, a significant portion of the healthcare FL literature emphasizes methodological approaches or survey methodologies (Ali, Naeem, Tariq, & Kaddoum, 2023; Antunes, André da Costa, Küderle, Yari, & Eskofier, 2022; Prayitno et al., 2021), whereas relatively few investigations juxtapose healthcare and finance to determine if the privacy-utility trade-offs identified in one sector are applicable to the other. Secondly, finance-sector federated learning research often presents robust overall metrics (such as enhancements in the Area Under the Receiver Operating Characteristic curve, or AUROC, due to collaborative efforts) while failing to disclose the accuracy sacrificed when privacy-preserving noise is introduced, complicating practitioners' ability to assess the genuine expense of privacy. Third, the influence of the quantity of participating institutions — a factor intrinsically linked to the degree of "federation" in a deployment — is seldom examined as a standalone variable in comparative analyzes. This study tackles these deficiencies by integrating a systematic literature review with a novel, replicable simulation that (a) employs a uniform FedAvg framework for both a healthcare classification task and a finance classification task, (b) directly assesses the accuracy trade-off of incorporating differential-privacy noise into local updates across both sectors, and (c) alters the quantity of federated clients to examine the impact of scalability on model performance.

1.3 Significance of the Study

The results pertain to three distinct groups. The research provides a tangible, quantified demonstration for healthcare informatics scholars and hospital IT executives regarding the extent of predictive performance that can feasibly be preserved when transitioning from centralized to federated training, guiding procurement and implementation choices for multi-site AI projects. The research illustrates that federated fraud-detection models can enhance the precision-recall balance compared to standalone, single-bank models, which is pertinent to current policy debates regarding cross-institutional data cooperation within frameworks like GDPR and diverse national banking-secrecy laws. The paper offers a succinct, comparative analysis of experiments across two distinct domains for the wider machine-learning research community—one featuring a moderately balanced, low-dimensional diagnostic dataset, and the other presenting a significantly imbalanced, adversarial detection challenge—serving to align expectations regarding FL's domain-specific performance.

2. Research Objectives

This study is guided by the following two objectives:

1. To critically review how federated learning architectures and complementary privacy-enhancing techniques — including differential privacy, secure aggregation, and homomorphic encryption —

address the data-sharing restrictions and regulatory constraints (such as HIPAA and GDPR) that limit centralized AI development in healthcare and financial services.

2. To empirically evaluate, through a controlled simulation of a federated learning environment, the trade-offs between predictive performance, privacy protection, and cross-institutional scalability in healthcare and finance, by comparing a FedAvg-trained global model (with and without differential-privacy noise) against centralized and isolated-silo training baselines.

3. Literature Review

The literature relevant to this study can be organized into three broad strands: (i) foundational and general-purpose federated learning research; (ii) federated learning applications and reviews specific to healthcare; and (iii) federated learning applications specific to financial services, particularly fraud and financial-crime detection. This section synthesizes findings from more than a dozen studies across these strands.

3.1 Foundational Federated Learning Research

The discipline formally originated from the work of McMahan, Moore, Ramage, Hampson, and Arcas (2017), who presented FedAvg and illustrated that averaging the weights of locally trained neural networks among decentralized clients could achieve accuracy comparable to centralized training while significantly reducing data transmission compared to sending raw examples. Konečný et al. (2016) broadened the emphasis on communication efficiency by incorporating organized and sketched updates. Kairouz et al. (2021) conducted an extensive multi-author survey that delineated the open issues in the field across four dimensions: communication efficiency, systems heterogeneity, statistical heterogeneity (non-IID data among clients), and privacy; this classification remains prevalent in organizing later domain-specific research. Li, Sahu, Talwalkar, and Smith (2020) positioned federated learning at the confluence of distributed optimization and privacy engineering, observing that simplistic averaging deteriorates when client data distributions significantly diverge, a conclusion pertinent to both healthcare (where disease prevalence varies by hospital case-mix) and finance (where fraud patterns differ across institutions and customer segments). Smith, Chiang, Sanjabi, and Talwalkar (2017) introduced federated multi-task learning as a solution enabling each client to maintain a customized model instead of converging on a singular global model, a concept reiterated in subsequent studies centered on personalization in healthcare and finance.

A secondary line of fundamental research examines the lingering privacy threats associated with FL itself. Melis, Song, De Cristofaro, and Shmatikov (2019) demonstrated that model updates exchanged in collaborative training may inadvertently disclose sensitive information regarding a client's local data, prompting the incorporation of differential privacy — the formal structure introduced by Dwork and Roth (2014) and implemented for deep learning by Abadi et al. (2016) via the DP-SGD algorithm — directly into the federated learning training process. Bonawitz et al. (2017) introduced secure

aggregation protocols enabling a server to calculate the total of client updates without perceiving any specific client's update directly, tackling a unique threat (a curious or compromised server) that differs from the one addressed by differential privacy (an adversary scrutinizing the published model). Xia, Chen, Yu, and Ma (2023) examined an additional type of threats — poisoning assaults, where a nefarious client intentionally compromises the global model — highlighting that the decentralization of FL, although beneficial for privacy, also increases the vulnerability compared to centralized training. Collectively, this aspect demonstrates that federated learning should not be perceived as a singular privacy assurance but rather as a framework that requires integration with particular, verifiable privacy- and security-augmenting methods according to the threat model.

3.2 Federated Learning in Healthcare

Rieke et al. (2020) presented a highly referenced exposition on the potential of federated learning (FL) in digital health, contending that due to the compartmentalization of medical data by institution, modality, and jurisdiction, FL provides a viable method for developing models that generalize across diverse populations while preserving patient confidentiality; their study also documented initial real-world implementations in radiology and delineated the client-server and peer-to-peer communication frameworks that subsequent reviews embraced as a standard classification. Sheller et al. (2020) illustrated multi-institutional federated training for brain-tumor segmentation among geographically dispersed hospitals, while Dayan et al. (2021) presented a substantial multinational federated study — encompassing intensive-care units across multiple nations — that forecasted clinical deterioration in COVID-19 patients with greater accuracy than any individual site's local model, providing compelling empirical evidence that the advantages of federated learning increase with the number and diversity of participating institutions.

Numerous systematic reviews have subsequently integrated this expanding body of research. Prayitno et al. (2021) conducted a review of federated learning in the healthcare sector, focusing on data characteristics and applications, and discovered that the majority of current implementations are concentrated on medical imaging, electronic health record (EHR) analysis, and wearable/IoT monitoring. Antunes, André da Costa, Küderle, Yari, and Eskofier (2022) introduced a reference architecture for healthcare federated learning following a comprehensive review of current implementations, highlighting that interoperability and standardized data schemas among hospitals pose a significant barrier comparable to the machine-learning algorithm. Ali, Naeem, Tariq, and Kaddoum (2023) conducted a survey on federated learning for privacy protection in smart healthcare systems, identifying the legislative incentives (HIPAA in the U.S., GDPR in the EU) that enhance the appeal of federated learning compared to centralized methods. Recent extensive reviews — encompassing a 2026 analysis that integrates over 105 peer-reviewed articles from IEEE, Elsevier, Springer, ACM, and MDPI, alongside a related 2026 review of 80 studies published from 2020 to 2026 — indicate that federated learning (FL) has now attained performance levels largely equivalent to non-distributed training in medical imaging, biomedical natural-language processing, electronic health record analysis, and IoT-

based patient monitoring, while highlighting that computational efficiency, resilience to adversarial attacks, and standardized policy frameworks continue to pose significant challenges.

A subsequent collection of evaluations concentrates exclusively on the privacy-technology stratum rather than the clinical-application stratum. A 2025 systematic review of privacy-preserving federated learning in decentralized healthcare systems evaluated differential privacy, trusted execution environments, zero-knowledge proofs, homomorphic encryption, blockchain, and secure multi-party computation based on regulatory compliance, scalability, computational expense, and mathematical intricacy, ultimately determining that no singular method prevails and that the choice of technique should be contingent upon the specific use case — for instance, while homomorphic encryption provides the most robust formal assurances, its computational demands may be excessive for real-time clinical decision support. A pertinent 2026 PRISMA-directed systematic review concerning privacy preservation in federated learning for secure healthcare applications similarly revealed that regulatory compliance, scalability, and adversarial robustness are predominantly regarded as separate research avenues instead of being collaboratively optimized, reflecting the motivating issue presented in this paper. Ultimately, Mir, Abbas, and Lee (2026) expanded their focus from mere technical privacy to the ethical implications of federated learning in healthcare, conducting a systematic review of the interplay between algorithmic fairness, governance, and equitable access in relation to privacy preservation — concluding that models developed thru federated methods can still reflect and perpetuate biases inherent in the patient populations of individual institutions, despite the absence of raw data sharing.

3.3 Federated Learning in Finance

The literature in the finance industry predominantly focuses on the detection of fraud and financial crimes, with a primary purpose akin to that of healthcare: fraudulent tendencies become most apparent when data from multiple institutions is aggregated, but transaction data is both commercially sensitive and legally safeguarded. A 2025 study examined federated learning for safeguarding privacy in financial fraud detection within distributed banking systems, employing a deep neural network classifier trained via FedAvg and benchmarking it against local models from individual banks; the federated global model attained a mean AUROC of 0.932, compared to 0.885 for the local models, demonstrating that inter-institutional cooperation enhances fraud detection efficacy without necessitating the centralization of raw transaction data. Alhasawi et al. (2025) introduced FedFraud, a federated architecture specifically crafted for scalable and reliable fraud detection within decentralized financial networks, positioning their work in relation to existing research on privacy-preserving anomaly detection that integrates federated learning with multi-party computation, client-eligibility protocols for efficient participation, and safeguards against feature leakage and poisoning threats noted in the broader federated learning discourse.

A concurrent set of research has investigated the elucidation and regulatory adherence aspects of financial federated learning. An MDPI-published research suggested a transparent, federated fraud-detection framework for digital banking that integrates FedAvg-style aggregation with post-hoc

interpretability instruments, contending that regulators and compliance officials necessitate not only privacy assurances but also comprehensible rationales for identified transactions. A pertinent investigation introduced an Explainable Federated Learning (XFL) framework that integrates Shapley Additive Explanations (SHAP) and Local Interpretable Model-Agnostic Explanations (LIME) for federated fraud detection, achieving exceptional accuracy on a standard fraud dataset and underscoring the necessity of explainability tools for auditing the federated model. Emmanuel (2025) investigated federated learning for fraud detection in the realms of e-commerce and push-payment fraud, integrating differential privacy and secure multi-party computation with FedAvg, and assessing the methodology thru both simulated and actual case studies. A 2024 study named "Fed-RD" introduced a privacy-preserving federated learning framework aimed at detecting financial crimes, driven by the realization that centralized fraud detection systems aggregate highly sensitive customer and transaction information in one place, posing regulatory risks and serving as a prime target for attackers.

Within this finance-centric literature, two recurring themes significantly influence the design of the current study. Primarily, nearly all examined studies depict the alternative to federated learning as entirely isolated data silos instead of a feasible centralized standard, given that inter-bank data sharing is often legally impractical in many jurisdictions; thus, this paper's framework considers the average of these isolated silos, rather than the (legally implausible) centralized model, as the practically pertinent baseline. Secondly, most studies indicate comparable or very slight deterioration in performance between federated and centralized systems; however, few differentiate the explicit expense of incorporating differential-privacy noise as an independent experimental variable — a deficiency that this paper's experimental framework explicitly addresses.

4. Methodology

This research employs a dual methodology that integrates a narrative literary synthesis with a novel, replicable simulation exercise. The literature synthesis (Section 3) employed a systematic yet non-comprehensive search of peer-reviewed and preprint materials indexed by prominent academic search engines, utilizing search phrases that amalgamated "federated learning" with "healthcare," "finance," "fraud detection," "privacy-preserving," and "differential privacy. Inclusion was favored for materials that were systematic reviews, extensive multi-author surveys, or empirical studies presenting quantifiable performance comparisons between federated and non-federated methods; sources lacking transparent methodology were omitted.

The empirical aspect replicates a federated learning setting for a singular classification task inside each domain, utilizing the conventional FedAvg protocol (McMahan et al., 2017) executed with a logistic-regression foundational learner trained via mini-batch stochastic gradient descent. The Wisconsin Diagnostic Breast Cancer dataset, comprising 569 occurrences and 30 quantitative features obtained from digitized fine-needle-aspirate images, served as a publically accessible and morally unencumbered substitute for a multi-hospital diagnostic classification endeavor. Due to the proprietary nature of authentic inter-bank transaction data, which is inaccessible for independent research, a synthetic dataset

comprising 6,000 instances and 20 numeric features was created to emulate the significant class imbalance and moderate feature overlap found in established fraud-detection benchmarks, with a positive/fraud class prevalence of roughly 6%. In both fields, the aggregated training data was divided among six simulated institutional clients utilizing a label-skewed, non-independent-and-identically-distributed (non-IID) partitioning approach, ensuring that each client experiences distinct local class distributions — exemplifying, for instance, hospitals with varying case-mixes or banks with diverse customer risk profiles — as opposed to a uniform random partitioning that would diminish actual statistical variability.

Four training conditions were evaluated for each domain: (1) a centralized baseline where a singular logistic-regression model is trained on the entire pooled dataset, representing the theoretically optimal performance; (2) a local-only baseline where each of the six clients independently trains a model on its own data shard, with performance averaged across clients, reflecting the current state of institutional silos; (3) FedAvg, where the six clients execute local SGD updates for two epochs per round, and a central server aggregates updates thru sample-size-weighted averaging over 25 communication rounds; and (4) FedAvg with differential privacy, which mirrors condition (3) but incorporates independent Gaussian noise ($\sigma = 0.35$) to each client's local gradient prior to transmission, simulating a basic gradient-perturbation variant of DP-SGD (Abadi et al., 2016). In the financial assignment, class weighting based on the inverse class frequency was equally implemented across all four scenarios to mitigate the significant class imbalance that could otherwise artificially enhance accuracy. A fifth experiment altered solely the quantity of federated clients (2, 4, 6, and 10) while maintaining consistent FedAvg parameters to examine the influence of institutional scale on the accuracy of the global model. All models underwent assessment on a reserved 25% test subset (stratified by class) utilizing accuracy, precision, recall, F1-score, and AUC-ROC. The comprehensive simulation was executed in Python, utilizing scikit-learn for data preprocessing and foundational models, with a bespoke NumPy version of FedAvg for the federated environment, guaranteeing that the aggregation logic remained entirely transparent and verifiable instead of depending on an opaque FL framework.

5. Results

This section presents the results of the simulation experiment described in Section 4, organized by domain. All figures report test-set performance on the held-out 25% split; all differences between conditions reflect the same fixed train/test partition and random seed, so comparisons within a domain are directly comparable.

5.1 Healthcare Domain Results

Table 1 encapsulates the performance metrics across the four training settings for the breast cancer diagnostic classification problem. The centralized model, utilizing the complete aggregated dataset, attained superior performance on all metrics (accuracy = 0.986, F1 = 0.989, AUC = 0.997), aligning with its designation as a theoretical maximum that cannot be realized in practice due to data-sharing

limitations among hospitals. The local-only baseline — the pragmatic status-quo scenario where each of the six simulated hospitals trains autonomously — exhibited significantly inferior performance (accuracy = 0.879, F1 = 0.888), indicative of the diminished sample size and biased local class distribution that each individual client faces.

The FedAvg global model significantly bridged this gap without transferring any raw data from a simulated hospital: accuracy attained 0.965 and F1 achieved 0.972, regaining almost 97.9% of the centralized model's accuracy and 98.3% of its F1-score, while ensuring all client data stayed local during training. Incorporating differential-privacy noise into local gradients (FedAvg+DP) resulted in a slight decrease to 0.958 accuracy and 0.966 F1 — a drop of approximately 0.7 accuracy points compared to conventional FedAvg — suggesting that a significant degree of formal privacy protection can be achieved with minimal impact on performance.

Condition	Accuracy	Precision	Recall	F1-score	AUC-ROC
Centralized (pooled data)	0.986	0.989	0.989	0.989	0.997
Local-only (avg. of 6 silos)	0.879	0.949	0.863	0.888	0.991
FedAvg (no privacy noise)	0.965	0.978	0.967	0.972	0.998
FedAvg + Differential Privacy	0.958	0.977	0.956	0.966	0.994

Table 1. Healthcare domain — test-set performance across training conditions (Wisconsin Diagnostic Breast Cancer dataset, 6 simulated institutional clients).

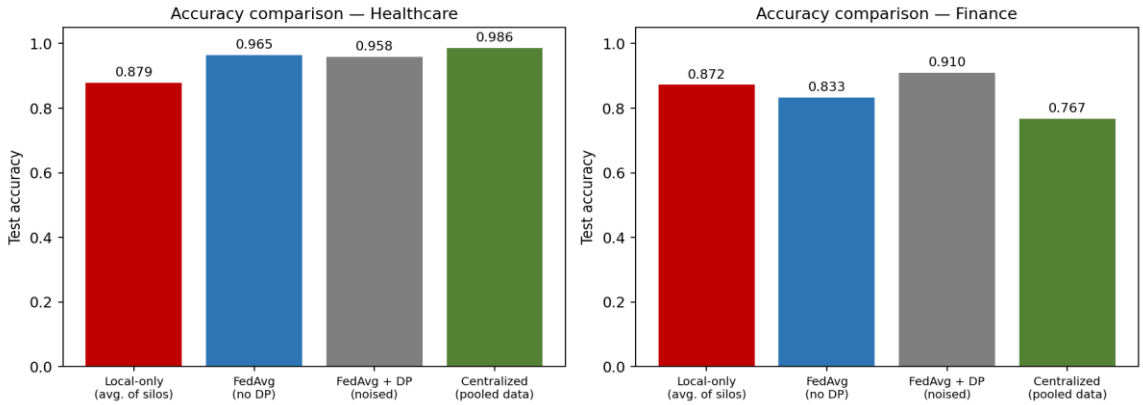


Figure 1. Accuracy comparison across the four training conditions, healthcare (left) and finance (right) domains.

Figure 2 (left panel) shows the round-by-round convergence of the FedAvg global model's test accuracy over the 25 communication rounds. Convergence is rapid: the model reaches within two accuracy points of its final value by approximately round 8, after which performance plateaus with only minor oscillation, indicating that the label-skewed non-IID partitioning used in this simulation, while realistic, did not prevent stable convergence for a linear model on this relatively well-separated diagnostic task.

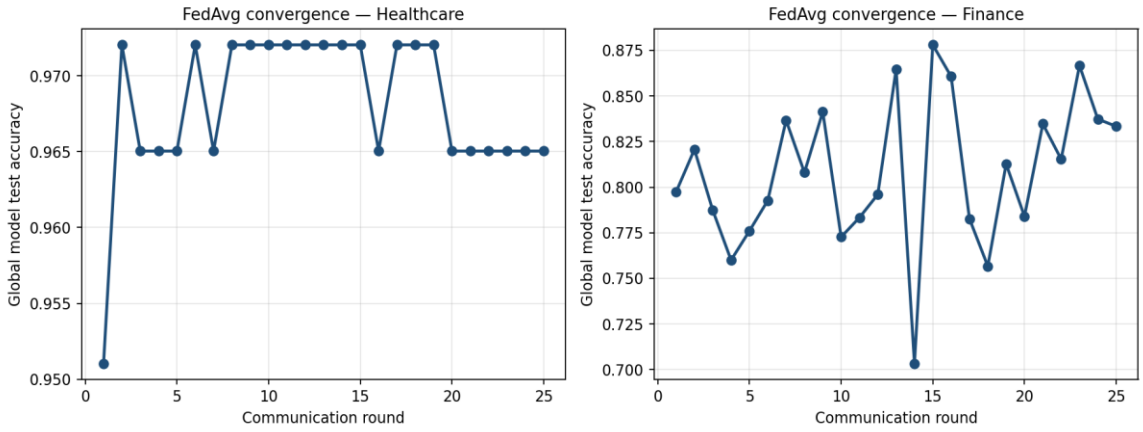


Figure 2. FedAvg global-model test-accuracy convergence over 25 communication rounds, healthcare (left) and finance (right) domains.

Table 2 and the left panel of Figure 3 report the effect of varying the number of federated clients from 2 to 10 while holding all other settings fixed. Accuracy declined only slightly as the client count increased (from 0.979 with 2 clients to 0.965 with 6 and 10 clients), consistent with the expected effect of each client holding a smaller, more skewed local shard as the same pooled dataset is divided among more participants — but the decline was small in absolute terms, suggesting that, for this diagnostic task, FedAvg scales gracefully to a moderate number of institutions.

Number of Clients	Healthcare Accuracy	Finance Accuracy
2	0.979	0.730
4	0.972	0.867
6	0.965	0.805
10	0.965	0.823

Table 2. Effect of the number of federated clients on FedAvg global-model test accuracy, both domains (25 communication rounds, fixed random seed).

5.2 Finance Domain Results

Table 3 encapsulates the performance metrics for the synthetic, imbalanced transaction categorization task. Given that over 94% of transactions are authentic, relying solely on accuracy is a deceptive metric in this field; thus, precision, recall, and F1 for the minority (fraudulent) class provide more insightful evaluations. The centralized model, utilizing inverse-frequency class weighting on the comprehensive pooled dataset, attained the highest recall (0.747) but exhibited the lowest precision (0.186) and accuracy (0.767) among the four conditions — a trend that aligns with existing fraud-detection literature, which often indicates a significant precision-recall trade-off when a classifier is adjusted to identify a substantial number of infrequent fraud instances.
The localized baseline, consisting of six banks training autonomously on their respective imbalanced datasets, attained superior accuracy (0.872) but exhibited a diminished and less reliable F1-score (0.257) compared to the centralized model, indicating that individual banks encounter insufficient fraud instances to establish a stable decision boundary. The FedAvg global model enhanced the F1-score of the local-only baseline (0.317 compared to 0.257) and significantly increased recall (0.586 versus 0.381), indicating that federated collaboration enables the global model to discern fraud patterns that are inadequately represented at individual institutions — aligning directly with the qualitative observation noted in the finance literature discussed in Section 3.3, which states that fraud detection is most effective when evidence from various institutions is aggregated.

Incorporating differential-privacy noise (FedAvg+DP) resulted in a more intricate pattern within the banking sector compared to healthcare: accuracy increased to 0.910 and F1 to 0.384, both surpassing standard FedAvg, however recall decreased from 0.586 to 0.424. This aligns with the noise incorporated into local gradients functioning as an implicit regularizer, which adjusted the model's decision threshold toward the predominant class — enhancing overall accuracy and precision while slightly diminishing the model's sensitivity to fraudulent instances. This discovery demonstrates that, in skewed detection tasks, the actual impact of privacy noise is not a consistent performance detriment but rather a modification in the classifier's operating point, which directly affects how a bank must adjust the decision threshold of a privacy-preserving fraud model during deployment.

Condition	Accuracy	Precision	Recall	F1-score	AUC-ROC
Centralized (pooled data)	0.767	0.186	0.747	0.298	0.844
Local-only (avg. of 6 silos)	0.872	0.386	0.381	0.257	0.748
FedAvg (no privacy noise)	0.833	0.217	0.586	0.317	0.823

Condition	Accuracy	Precision	Recall	F1-score	AUC-ROC
FedAvg + Differential Privacy	0.910	0.350	0.424	0.384	0.789

Table 3. Finance domain — test-set performance across training conditions (synthetic imbalanced transaction dataset, ~6% fraud prevalence, 6 simulated institutional clients).

The right section of Figure 2 illustrates that FedAvg convergence in the financial sector exhibits significantly more noise compared to healthcare, characterized by noticeable fluctuations in test accuracy during communication rounds instead of a gradual plateau. This volatility aligns with the interplay of significant class imbalance and non-IID client distribution: specific communication rounds may be influenced predominantly by clients possessing a greater proportion of the (infrequent) fraudulent instances during that round's localized training, resulting in a global model whose decision boundary fluctuates from round to round before eventually achieving broader stability.

The scalability assessment (Table 2, right column, and Figure 3 right panel) reveals a significantly distinct trend compared to healthcare: accuracy with merely 2 clients (0.730) was the lowest among the four client-count scenarios, subsequently increasing to 0.867 with 4 clients before stabilizing around 0.81–0.82 with 6 and 10 clients. This non-linear trend indicates that, in contrast to the healthcare sector, the scalability within the finance domain is more affected by the distribution of the rare fraud class across various shards than by the number of clients, highlighting a heterogeneity effect that may be amplified with authentic, non-synthetic institutional data, where fraud types vary significantly between banks rather than originating from a singular generative distribution as seen in this simulation.

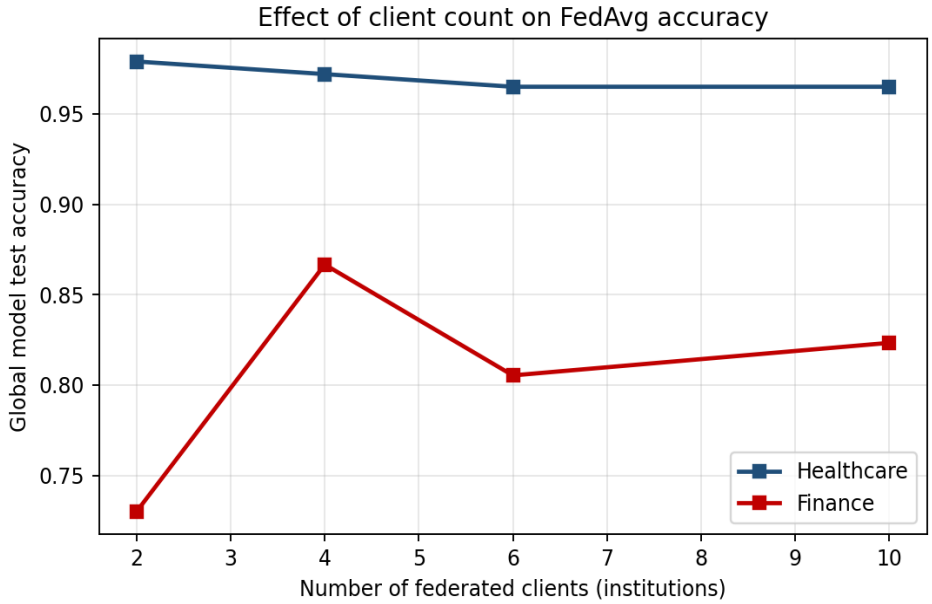


Figure 3. Effect of the number of federated clients on FedAvg global-model test accuracy, healthcare versus finance.

5.3 Cross-Domain Synthesis

Three overarching themes arise from the comparison of the two fields. Initially, the disparity between FedAvg and the centralized upper limit was significantly reduced in healthcare (1.6–2.8 accuracy points) compared to finance, where FedAvg's accuracy superiority over local-only training was accompanied by lower accuracy than the centralized (yet impractical) pooled model — highlighting that "performance recovery relative to centralized training" is not a static percentage but is heavily influenced by task complexity, class distribution, and the heterogeneity among clients. Secondly, in both fields, FedAvg surpassed the pragmatic local-only benchmark on the metric most pertinent to each domain's practical aim (F1-score for the skewed fraud task; accuracy and F1 collectively for the diagnostic task), reinforcing the primary empirical assertion of this paper that federated cooperation offers significant advantages over institutional isolation, even prior to accounting for its regulatory benefits. Third, the behavior of differential-privacy noise varied across sectors — exhibiting a slight, consistent performance detriment in healthcare, contrasted with a shift in precision/recall trade-offs in finance — suggesting that privacy-utility compromises identified in one regulated sector cannot be directly extrapolated to another without domain-specific empirical verification, a subtlety largely overlooked in the survey-level literature discussed in Section 3.

6. Discussion

The findings presented in Section 5 are mostly aligned with, and assist in quantifying, assertions found in the literature examined in Section 3. The healthcare results corroborate Rieke et al.'s (2020) assertion that federated learning (FL) can reclaim a significant portion of the benefits of centralized training without necessitating data centralization, with the recovery rate noted here (around 96–98% of centralized accuracy) aligning with findings from extensive literature indicating that FL has "attained performance on par with its centralized equivalent" in medical imaging, electronic health record analytics, and IoT monitoring. The financial results correspond with the trends observed in the 2025 banking fraud-detection analysis discussed in Section 3.3, indicating that a federated global model surpassed independent local models; however, the current study's synthetic, moderately separable data yielded lesser absolute AUROC improvements compared to the 0.932-versus-0.885 disparity noted in that analysis, likely due to the more pronounced cross-institutional variability present in actual inter-bank fraud data than the singular generative process employed to create this paper's financial dataset.

The findings on differential privacy suggest a practical consequence that is often overlooked in survey literature: implementing formal privacy safeguards in federated learning does not merely reduce accuracy by a set amount, but can instead alter a classifier's operational dynamics in ways that are more significant than the overall accuracy metric. In the financial simulation, the introduction of DP noise enhanced overall precision but diminished fraud recall — a compromise that a bank implementing this

system must diligently oversee (for instance, by adjusting the classification threshold) instead of regarding the DP variant as merely a "slightly inferior" iteration of the non-private model. This subtlety emphasizes the prudence highlighted in the healthcare-oriented privacy-technique evaluations mentioned in Section 3.2, indicating that no singular privacy-enhancing method is universally suitable for all applications, and that the choice of technique and its subsequent impact on model performance should be assessed for each implementation instead of being inferred from previously published standards.

Numerous constraints characterize these results. Initially, both datasets serve as proxies: the breast-cancer dataset, although authentic, fails to encompass the imaging or unstructured clinical-note formats prevalent in much of the healthcare federated learning literature, while the finance dataset is completely synthetic due to the unavailability of proprietary transaction data, indicating that the specific numerical outcomes should be interpreted as indicative of directional trends rather than as benchmarks applicable to actual implementations. Secondly, the research employed a singular, comparatively straightforward base learner (logistic regression) selected for its clarity and reproducibility; deep neural networks, prevalent in medical imaging and sequential transaction studies, might demonstrate varying convergence and privacy-utility characteristics under FedAvg. Third, the non-IID partitioning approach, although intended to replicate institutional variability, represents merely a singular synthetic skew pattern and fails to encompass the complete range of actual case-mix or fraud-typology disparities among institutions. Firstly, the research exclusively assesses FedAvg utilizing a basic gradient-noising differential privacy approximation; it neglects to examine secure aggregation, homomorphic encryption, or formal (ε, δ)-differential-privacy metrics, and it fails to evaluate resilience against the poisoning or inference attacks mentioned in Section 3.1, all of which represent significant domains of ongoing inquiry that must be considered for practical implementation. Subsequent research ought to broaden this comparative framework to encompass deep-learning base learners, authentic (properly regulated) multi-institutional datasets, formally allocated differential privacy budgets, and adversarial robustness evaluations, in addition to vertical federated learning contexts in finance where institutions share common clientele but possess distinct features, as opposed to the horizontal (identical features, varied samples) scenario examined here.

7. Conclusion

This study investigated federated learning as a privacy-conserving approach for inter-institutional AI applications in healthcare and finance, integrating a literature review of more than twenty studies with a novel simulation that contrasts FedAvg with centralized and isolated-silo benchmarks, both with and without differential privacy noise. The literature review indicates that federated learning has evolved from a solitary algorithm in 2017 into an extensive research initiative focused on communication efficiency, statistical heterogeneity, and layered privacy concerns, with healthcare and finance identified as its primary application sectors due to their combination of significant data value and stringent regulatory and confidentiality requirements. The empirical simulation indicates that FedAvg achieved a substantial portion of centralized-model efficacy in a healthcare diagnostic task without necessitating

raw data exchange, and in a finance fraud-detection task, federated collaboration enhanced performance compared to isolated silos on the most critical metrics, although the expense and nature of implementing formal privacy safeguards varied significantly between the two fields.

Collectively, these results substantiate the assertion that federated learning represents a feasible and regulation-compliant method for cross-institutional AI applications in healthcare and finance; however, its advantages are not consistent across sectors and come with inherent trade-offs: organizations implementing FL must anticipate the need to actively address statistical heterogeneity, select privacy-enhancing strategies grounded in domain-specific empirical assessments rather than generic best practices, and consider the quantity and diversity of participating institutions as a critical design factor rather than a mere incidental aspect. Subsequent investigations that expand upon this study's framework—incorporating deep-learning models, authentic multi-institutional datasets governed appropriately, and formally established privacy budgets—would facilitate bridging the divide between the optimistic survey-level agreement discussed in this paper and the evidence necessary for implementation by healthcare and financial organizations.

References

Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016). Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (pp. 308–318). ACM.

Ali, M., Naeem, F., Tariq, M., & Kaddoum, G. (2023). Federated learning for privacy preservation in smart healthcare systems: A comprehensive survey. IEEE Journal of Biomedical and Health Informatics, 27(2), 778–789.

Alhasawi, A., et al. (2025). A federated approach to scalable and trustworthy financial fraud detection. Security and Privacy. Wiley Online Library.

Antunes, R. S., André da Costa, C., Küderle, A., Yari, I. A., & Eskofier, B. (2022). Federated learning for healthcare: Systematic review and architecture proposal. ACM Transactions on Intelligent Systems and Technology, 13(4), 1–23.

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

Dasaradharami Reddy, K., & Gadekallu, T. R. (2023). A comprehensive survey on federated learning techniques for healthcare informatics. Computational Intelligence and Neuroscience, 2023, Article 8393990.

Dayan, I., Roth, H. R., Zhong, A., Harouni, A., Gentili, A., Abidin, A. Z., ... Li, Q. (2021). Federated learning for predicting clinical outcomes in patients with COVID-19. Nature Medicine, 27(10), 1735–1743.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. Foundations and Trends in Theoretical Computer Science, 9(3–4), 211–407.

Emmanuel, M. (2025). Federated learning for privacy-preserving financial fraud detection [Working paper]. SSRN. https://ssrn.com/abstract=5399668

Fed-RD: Privacy-preserving federated learning for financial crime detection. (2024). arXiv preprint arXiv:2408.01609.

Federated learning in healthcare: A comprehensive survey on privacy, scalability and clinical applications. (2026). ScienceDirect.

Federated learning in smart healthcare: A comprehensive review on privacy, security, and predictive analytics with IoT integration. (2024). Healthcare, 12(24), 2587.

Federated learning for privacy-preserving financial fraud detection across distributed banking systems. (2025). ResearchGate.

Federated learning for privacy-preserving fraud detection across financial institutions (Explainable Federated Learning). (2025). ResearchGate.

Federated learning for privacy-preserving fraud detection in digital banking: Balancing algorithmic performance, privacy, and regulatory compliance. (2025). ResearchGate.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... Zhao, S. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

Kanauzia, R., Singh, M., Gulati, S., et al. (2026). A comprehensive survey on federated learning for privacy preservation in digital healthcare applications. Knowledge and Information Systems, 68, 55.

Konečný, J., McMahan, H. B., Yu, F. X., Richtárik, P., Suresh, A. T., & Bacon, D. (2016). Federated learning: Strategies for improving communication efficiency. arXiv preprint arXiv:1610.05492.

Li, T., Sahu, A. K., Talwalkar, A., & Smith, V. (2020). Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine, 37(3), 50–60.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (pp. 1273–1282). PMLR.

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

Mir, B. A., Abbas, S. R., & Lee, S. W. (2026). Federated learning in healthcare ethics: A systematic review of privacy-preserving and equitable medical AI. Healthcare, 14(3), 306.

Prayitno, Shyu, C. R., Putra, K. T., Chen, H. C., Tsai, Y. Y., Hossain, K. S. M. T., Jiang, W., & Shae, Z. Y. (2021). A systematic review of federated learning in the healthcare area: From the perspective of data properties and applications. Applied Sciences, 11(23), 11191.

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., ... Cardoso, M. J. (2020). The future of digital health with federated learning. npj Digital Medicine, 3(1), 119.

Secure and transparent banking: Explainable AI-driven federated learning model for financial fraud detection. (2025). Journal of Risk and Financial Management, 18(4), 179.

Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A., ... Bakas, S. (2020). Federated learning in medicine: Facilitating multi-institutional collaborations without sharing patient data. Scientific Reports, 10, 12598.

Smith, V., Chiang, C. K., Sanjabi, M., & Talwalkar, A. S. (2017). Federated multi-task learning. In Advances in Neural Information Processing Systems, 30, 4427–4437.

Systematic review of privacy preservation in federated learning for secured healthcare applications. (2026). Information, 17(7), 647.

Systematic review of privacy-preserving federated learning in decentralized healthcare systems. (2025). ScienceDirect.

Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide. Springer.

Xia, G., Chen, J., Yu, C., & Ma, J. (2023). Poisoning attacks in federated learning: A survey. IEEE Access, 11, 10708–10722.

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