10.48047/jocaaa.2023.31.04.81

Federated Learning and Privacy-Preserving AI for Cross-Institutional Financial Fraud Detection

Gopinathan Rathinavelu
GTSS Inc, Data Architect, Arizona, USA
Email ID: gopinathan.r@gmail.com

Abstract

Financial fraud increasingly operates across institutional boundaries, while conventional fraud-detection systems remain constrained by fragmented datasets, privacy requirements, data-sovereignty obligations, and competitive barriers to information sharing. This paper develops a privacy-preserving federated learning framework that enables multiple financial institutions to collaboratively train fraud-detection models without transferring raw transaction records. The framework integrates federated optimization, secure aggregation, differential privacy, encrypted model-update exchange, and robust aggregation to protect sensitive information against inference, poisoning, and unauthorized reconstruction attacks. Particular attention is given to non-independent and non-identically distributed transaction data, extreme class imbalance, institutional heterogeneity, communication overhead, and model drift. The proposed architecture supports local model training while generating a jointly optimized fraud-detection model from protected updates. The study establishes an analytical framework for evaluating detection performance, privacy leakage, communication efficiency, robustness, and operational compliance simultaneously. The resulting approach provides a foundation for collaborative financial intelligence in which institutions can obtain cross-organizational learning benefits while retaining control over their underlying customer and transaction data.

Keywords: *Federated Learning, Financial Fraud Detection, Privacy-Preserving Machine Learning, Secure Aggregation, Differential Privacy, Collaborative Financial Intelligence*

1. Introduction

The rapid expansion of digital banking, electronic payments, mobile financial services, and interconnected fintech platforms has substantially increased the scale and complexity of financial fraud. Fraudulent activities increasingly exploit multiple accounts, payment channels, merchants, devices, and financial institutions, creating patterns that may remain invisible when each institution analyzes only its own transaction records. Traditional fraud-detection architectures generally depend on centralized access to large historical datasets, where machine-learning models are trained by combining transaction attributes, behavioral indicators, authentication information, and historical fraud labels. Although centralized learning can provide broad statistical coverage, transferring sensitive financial records between institutions creates significant privacy, confidentiality, governance, and data-sovereignty concerns. Recent research identifies federated learning as an important architectural alternative because institutions can collaboratively improve a shared model while retaining transaction data within their respective environments [1], [2].

Federated learning (FL) changes the conventional data-sharing paradigm by moving model computation toward the data rather than moving data toward a centralized analytical repository. In a cross-institutional financial environment, each participating bank, payment processor, or fintech organization can train a local fraud-detection model using its proprietary transaction dataset. Instead of transmitting individual customer records, the institution communicates selected model parameters or updates to an aggregation mechanism, which combines

10.48047/jocaaa.2023.31.04.81

information from participating institutions to produce an updated global model. The fundamental federated optimization concept was established by McMahan et al., who demonstrated communication-efficient distributed model training through iterative local optimization and parameter aggregation [3]. This paradigm is particularly relevant to financial fraud detection because the underlying transaction data can remain institutionally controlled while statistical knowledge derived from multiple institutions is incorporated into a collaborative model.

However, federated learning should not be considered equivalent to complete privacy protection. Model parameters and gradients can themselves contain information about local training data and may become targets for inference, reconstruction, or membership attacks. Consequently, a privacy-preserving financial FL architecture requires additional protection mechanisms beyond decentralized training. Secure aggregation can prevent an aggregation server from directly observing individual institutional updates, while differential privacy can introduce mathematically controlled uncertainty into model updates and reduce the possibility of extracting information about individual records [4], [5]. Bonawitz et al. established practical secure aggregation mechanisms for privacy-preserving distributed learning, demonstrating how individual client updates can be concealed while still allowing their aggregate to be computed [4]. Similarly, differential privacy provides a formal framework for limiting the influence of individual records on model outputs and has been incorporated into machine-learning systems through techniques such as gradient clipping and calibrated noise addition [5].

Financial fraud detection introduces additional methodological difficulties because transaction data are highly heterogeneous and fraudulent transactions normally constitute only a small fraction of the overall transaction population. Different institutions may operate in different geographical regions, serve different customer groups, support different payment instruments, and experience different fraud typologies. Consequently, the statistical distribution of transaction features and fraud labels can differ substantially among participating institutions. This produces a non-independent and non-identically distributed (non-IID) federated environment, which can destabilize model convergence and reduce the effectiveness of conventional federated averaging. Research on non-IID federated learning has demonstrated that statistical heterogeneity can substantially affect convergence, model generalization, and participant contribution [6], [7]. In financial applications, this problem is intensified by institution-specific fraud patterns, class imbalance, delayed fraud labels, changing attack strategies, and differences in transaction volume.

The privacy challenge is therefore multidimensional. Protecting raw transaction data does not necessarily protect information contained in gradients, model parameters, intermediate representations, or model outputs. Furthermore, a malicious participant may attempt model poisoning by submitting manipulated updates designed to degrade the global model or introduce targeted vulnerabilities. Conversely, an honest-but-curious aggregator may attempt to infer sensitive information from individual updates. A robust cross-institutional architecture consequently requires coordinated mechanisms for confidentiality, privacy, authentication, secure aggregation, adversarial robustness, and model governance. Recent financial FL research increasingly treats privacy as a system-level property involving multiple complementary mechanisms rather than as a consequence of decentralizing the training data alone [1], [8].

Another important issue is the severe class imbalance inherent in fraud detection. A model trained primarily on legitimate transactions can achieve high overall accuracy while performing inadequately on the minority fraud class. Therefore, evaluation must extend beyond accuracy

10.48047/jocaaa.2023.31.04.81

to include precision, recall, F1-score, receiver operating characteristic area under the curve (ROC-AUC), precision-recall area under the curve (PR-AUC), false-positive rate, and fraud-detection sensitivity. The integration of federated learning with fraud detection has consequently attracted attention as a means of improving access to diverse learning patterns without requiring institutions to exchange raw transaction records [8], [9].

Explainability is also increasingly relevant because financial institutions require operationally interpretable fraud alerts and defensible model decisions. A highly accurate collaborative model may nevertheless create implementation difficulties if fraud analysts cannot understand the principal factors associated with a prediction. Research combining federated learning and explainable artificial intelligence has therefore emphasized transparency alongside predictive performance and privacy protection [9]. For financial institutions, explainability is particularly important when model outputs influence transaction blocking, account restrictions, enhanced due diligence, or other operational decisions.

The central research problem addressed in this paper is therefore the design of a collaborative fraud-detection mechanism that simultaneously addresses predictive performance, privacy preservation, heterogeneous institutional data, communication efficiency, and security. The proposed research direction considers federated learning as the collaborative foundation and incorporates secure aggregation, differential privacy, robust aggregation, and appropriate evaluation mechanisms to establish a privacy-preserving cross-institutional fraud-detection framework. The objective is not merely to prevent raw-data exchange but to develop an integrated architecture in which financial institutions can obtain collective learning benefits while retaining institutional control over sensitive transaction information.

2. Literature Review

2.1 Federated Learning as a Privacy-Preserving Learning Paradigm

Federated learning was introduced as a distributed machine-learning paradigm in which model training occurs across decentralized data sources while a coordinating mechanism aggregates locally generated model updates [3]. Unlike conventional centralized learning, FL does not require all participating datasets to be transferred to a common repository. This characteristic makes it suitable for domains in which data are highly sensitive, institutionally controlled, or legally restricted. Kairouz et al. subsequently provided a comprehensive analysis of federated learning challenges, including statistical heterogeneity, system heterogeneity, communication efficiency, privacy, security, and robustness [1]. In the financial sector, these challenges become particularly important because participating organizations generally possess heterogeneous datasets and operate under different institutional governance structures.

Cross-silo federated learning is particularly relevant to financial institutions because the participating entities are generally known organizations rather than millions of individual consumer devices. Banks, insurance companies, payment processors, and fintech organizations can therefore participate as relatively stable FL clients. Recent research on privacy-preserving FL for financial applications identifies cross-silo learning, federated optimization, secure aggregation, differential privacy, and hybrid privacy-enhancing technologies as important components of financial collaborative intelligence [1]. This creates a technical foundation for collaborative fraud detection without requiring centralized transaction-data repositories.

2.2 Machine Learning for Financial Fraud Detection

Machine-learning-based fraud detection has evolved from conventional statistical and rule-based systems toward supervised learning, anomaly detection, ensemble learning, deep

10.48047/jocaaa.2023.31.04.81

learning, and graph-based approaches. However, fraud detection remains fundamentally challenging because fraudulent transactions are rare relative to legitimate transactions and because fraud strategies evolve dynamically. Class imbalance can cause conventional classification models to prioritize the majority class, resulting in high aggregate accuracy but inadequate identification of fraudulent events. Recent studies specifically addressing federated fraud detection have incorporated data-balancing techniques to improve the effectiveness of distributed models under these conditions [8].

A further limitation of institution-specific fraud detection is restricted observational coverage. An individual bank may observe only one segment of a fraudulent campaign, while another institution may encounter a different stage of the same activity. Cross-institutional learning can potentially expose shared statistical characteristics without requiring the direct exchange of individual transaction records. Research on federated fraud detection has consequently investigated collaborative models capable of exploiting distributed transaction intelligence while preserving institutional data boundaries [8], [10].

The literature also demonstrates growing interest in explainable federated fraud detection. The combination of FL and explainable AI has been proposed to address the dual requirements of privacy-preserving collaborative learning and interpretable fraud decisions [9]. This direction is particularly significant in financial environments where analysts must investigate alerts and institutions must maintain auditable decision processes.

2.3 Secure Aggregation and Differential Privacy

Federated learning alone does not guarantee protection against information leakage. If individual model updates are visible to an aggregation server, information about the underlying training records may potentially be inferred. Secure aggregation addresses this issue by enabling the server to obtain an aggregate of participant updates without directly accessing each individual contribution. Bonawitz et al. developed a practical secure aggregation protocol specifically for privacy-preserving machine learning, establishing an important cryptographic foundation for federated systems [4].

Differential privacy provides a complementary privacy mechanism by limiting the contribution that an individual record can make to a model update. Abadi et al. demonstrated the use of differentially private stochastic gradient methods for deep learning through gradient clipping and calibrated noise addition [5]. In financial fraud detection, differential privacy can provide a formal privacy guarantee, but excessive noise may reduce sensitivity to rare fraud patterns. Consequently, privacy protection must be balanced against fraud-detection utility.

Recent research in financial privacy-preserving computation indicates that no single privacy-enhancing mechanism simultaneously optimizes privacy, predictive performance, computational efficiency, and compliance requirements. Federated learning, secure aggregation, differential privacy, homomorphic encryption, and secure multi-party computation therefore increasingly appear as complementary components within hybrid architectures rather than mutually exclusive alternatives [2]. This observation is directly relevant to the proposed framework because protecting financial transaction information requires security at multiple stages of the collaborative learning lifecycle.

2.4 Non-IID Data and Institutional Heterogeneity

One of the principal limitations of conventional federated optimization is the assumption that local datasets can contribute reasonably compatible statistical information. In real-world financial networks, this assumption is rarely satisfied. Institutions differ in transaction volumes,

10.48047/jocaaa.2023.31.04.81

customer demographics, geographical coverage, payment channels, merchant categories, fraud prevalence, and detection policies. Consequently, local datasets may exhibit significant feature distribution and label distribution differences.

Research on non-IID federated learning has demonstrated that statistical heterogeneity can negatively affect convergence and global-model performance [6], [7]. The problem becomes especially severe in fraud detection because fraudulent transactions are sparse and their distribution may be strongly institution-specific. A fraud pattern that appears frequently at one institution may be almost absent at another. Uniform parameter averaging can therefore cause important minority patterns to become diluted during global aggregation.

This challenge has motivated approaches involving personalized federated learning, clustered federated learning, adaptive aggregation, and optimization techniques designed for heterogeneous clients. Recent financial FL research similarly emphasizes non-IID robustness as a major requirement for practical cross-institution deployment [1], [10]. The literature consequently suggests that a successful financial FL framework should account for both global knowledge transfer and institution-specific characteristics rather than assuming identical local distributions.

2.5 Security Threats in Federated Fraud Detection

Federated financial systems face threats from both external attackers and participating clients. Model poisoning represents one important threat in which a malicious participant submits deliberately manipulated model updates to degrade or manipulate the global model. Inference attacks represent another threat category in which an adversary attempts to recover sensitive information from model parameters, gradients, or outputs. Secure aggregation can reduce exposure of individual updates, while differential privacy can limit information leakage from training processes [1], [4], [5].

The literature increasingly recognizes that privacy and integrity are separate requirements. A system may conceal individual model updates while remaining vulnerable to malicious contributions. Therefore, robust aggregation, participant authentication, anomaly detection, update validation, and model monitoring must complement privacy mechanisms. Recent cross-institutional fraud-detection architectures explicitly identify poisoning, inference attacks, non-IID data, secure aggregation, and differential privacy as interconnected design considerations [10].

Emerging research has also investigated hierarchical and adaptive federated architectures to address heterogeneous fraud distributions and privacy–utility trade-offs. Recent work on hierarchical privacy-preserving fraud detection reports the use of institution clustering, knowledge transfer, and adaptive differential-privacy mechanisms to improve performance under heterogeneous fraud distributions [11]. Such approaches indicate a shift from simple FedAvg-style architectures toward more specialized federated systems designed specifically for financial fraud characteristics.

2.6 Critical Research Gap

The literature establishes that federated learning can reduce the need for centralized financial-data sharing, while secure aggregation and differential privacy can provide additional protection against information leakage [1], [4], [5]. Nevertheless, several research gaps remain. First, many studies evaluate federated fraud detection under relatively controlled datasets rather than realistic cross-institution environments with substantial differences in transaction distributions. Second, privacy protection is frequently evaluated separately from fraud-

10.48047/jocaaa.2023.31.04.81

detection performance, making it difficult to quantify the practical privacy–utility trade-off. Third, communication overhead and computational costs receive less attention than predictive accuracy despite their importance in multi-institution deployment. Fourth, model poisoning and inference attacks require simultaneous consideration because confidentiality and model integrity represent different security objectives. Finally, governance, explainability, auditability, model drift, and operational deployment need to be integrated into the technical architecture rather than treated as independent post-deployment considerations.

Recent reviews of privacy-preserving federated learning in finance similarly identify non-IID robustness, communication costs, privacy–utility trade-offs, governance complexity, benchmarking inconsistencies, and limited real-world cross-institutional deployment as persistent challenges [1], [2]. The proposed study therefore focuses on an integrated privacy-preserving federated framework that combines collaborative fraud learning with secure aggregation, differential privacy, robustness mechanisms, heterogeneous-data handling, and multidimensional performance evaluation. This integrated perspective provides the foundation for the subsequent methodological and analytical sections of the paper.

3. Proposed Privacy-Preserving Federated Learning Framework

3.1 Overall System Architecture

The proposed framework is designed as a cross-institutional, cross-silo federated learning architecture in which participating banks, payment service providers, fintech organizations, and other authorized financial institutions retain transaction data within their respective secure environments. The architecture consists of six principal layers: **local data layer, preprocessing and feature-engineering layer, local fraud-learning layer, privacy-protection layer, federated aggregation layer, and global fraud-intelligence layer**. This architecture follows the emerging direction of financial federated learning in which raw transaction records remain within institutional boundaries while protected model information is exchanged for collaborative learning [1], [2].

Each institution I_k, where $k = 1,2, \dots, K$, maintains a local transaction dataset

$$D_k = \{(x_i, y_i)\}_{i=1}^{N_k}$$

where x_i represents the transaction feature vector, $y_i \in \{0,1\}$ represents the legitimate or fraudulent class, and N_k denotes the number of local transactions at institution k.

The complete distributed dataset is therefore represented as

$$D = \bigcup_{k=1}^{K} D_k$$

but D is never physically transferred to a centralized server.

A representative configuration containing **10 participating institutions** can be used to model the proposed cross-institution environment.

Table 1. Representative Characteristics of Participating Financial Institutions

Institution	Primary Transaction Type	Relative Data Volume	Fraud Prevalence	Local Data Distribution
FI-1	Retail banking	High	0.42%	Consumer dominated
FI-2	Credit cards	Very High	0.58%	Card dominated
FI-3	Digital payments	Very High	0.71%	Mobile dominated

10.48047/jocaaa.2023.31.04.81

Institution	Primary Transaction Type	Relative Data Volume	Fraud Prevalence	Local Data Distribution
FI-4	Corporate banking	Medium	0.31%	Business dominated
FI-5	Fintech payments	High	0.84%	Digital dominated
FI-6	E-commerce payments	High	0.93%	Merchant dominated
FI-7	Regional banking	Medium	0.36%	Regional dominated
FI-8	Online banking	High	0.63%	Internet dominated
FI-9	Payment gateway	Very High	0.76%	Merchant dominated
FI-10	Mixed financial services	Medium	0.49%	Mixed distribution

The substantial variation in fraud prevalence illustrates the non-IID nature of the proposed environment. Such heterogeneity is a major concern in federated financial learning because local models may learn institution-specific fraud characteristics rather than an identical representation of fraudulent behavior [1], [3].

3.2 Local Data Processing and Feature Engineering

Raw transaction information remains inside the institution's protected infrastructure. Local preprocessing begins with removal of duplicate records, correction of inconsistent fields, treatment of missing values, categorical encoding, normalization, and temporal feature extraction.

For transaction i, the feature vector can be represented as

$$x_i = [A_i, T_i, L_i, M_i, C_i, D_i, B_i, H_i, R_i]$$

where A_i represents transaction amount, T_i transaction time, L_i geographical or location characteristics, M_i merchant information, C_i channel information, D_i device characteristics, B_i behavioral indicators, H_i historical transaction statistics, and R_i risk-related attributes.

Continuous variables are normalized using

$$x_i' = \frac{x_i - \mu}{\sigma}$$

where μ and σ denote the local mean and standard deviation.

For highly skewed financial variables such as transaction amount, logarithmic transformation can be applied:

$$A_i' = \log(1 + A_i)$$

Temporal behavior can additionally be represented through variables such as transaction frequency, time since previous transaction, number of transactions within a rolling window, and deviation from the customer's historical spending profile.

Because fraudulent transactions constitute a minority class, class-balanced local training is required. A weighted binary cross-entropy objective can be defined as

$$L = -\frac{1}{N}\sum_{i=1}^{N}[w_1 y_i \log(\hat{y}_i) + w_0(1 - y_i)\log(1 - \hat{y}_i)]$$

where $w_1 > w_0$ provides greater importance to fraudulent transactions.

10.48047/jocaaa.2023.31.04.81

This approach is preferable to optimizing ordinary accuracy because a classifier could otherwise achieve apparently high accuracy by predominantly predicting the majority legitimate class. Recent federated fraud-detection studies have similarly emphasized data balancing and non-IID conditions as important factors in financial model development [8].

3.3 Local Fraud-Detection Model

Each institution independently trains a local fraud-detection model using its private dataset. A hybrid neural architecture can be adopted in which dense layers capture nonlinear transaction relationships while temporal or sequential layers capture behavioral evolution.

For institution k, the local model is denoted as

$$f_{\theta_k}(x)$$

where θ_k represents the local model parameters.

The local optimization objective is

$$\theta_k^{(t+1)} = \theta_k^{(t)} - \eta \nabla L_k(\theta_k^{(t)})$$

where η is the learning rate and L_k represents the local fraud-detection loss.

For a mini-batch B_k, the gradient is

$$g_k = \frac{1}{\mid B_k \mid} \sum_{i \in B_k} \nabla_\theta \, L(f_\theta(x_i), y_i)$$

The model can generate a fraud probability

$$P(y = 1 \mid x) = \sigma(z)$$

where

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

and z is the final model logit.

A transaction is classified as fraudulent when

$$P(y = 1 \mid x) \geq \tau$$

where τ is the operational fraud-detection threshold.

The threshold should not necessarily be fixed at 0.5 because financial fraud detection is highly asymmetric. A lower threshold can improve recall but may increase false positives, whereas a higher threshold can reduce unnecessary alerts but potentially miss fraudulent transactions.

3.4 Federated Optimization and Global Model Formation

After local training, each institution produces a model update rather than transmitting transaction-level information. The standard weighted federated aggregation process can be expressed as

10.48047/jocaaa.2023.31.04.81

$$\theta^{(t+1)} = \sum_{k=1}^{K} \frac{N_k}{\sum_{j=1}^{K} N_j} \theta_k^{(t+1)}$$

where the weighting factor

$$w_k = \frac{N_k}{\sum_{j=1}^{K} N_j}$$

reflects the relative contribution of institution k.

However, direct FedAvg may not be sufficient under strongly heterogeneous financial datasets. A trust-adjusted aggregation mechanism is therefore incorporated:

$$\theta^{(t+1)} = \frac{\sum_{k=1}^{K} w_k \, q_k \theta_k^{(t+1)}}{\sum_{k=1}^{K} w_k \, q_k}$$

where q_k denotes the reliability score of institution k.

The reliability score can incorporate:

- historical update consistency,
  validation performance,
  update magnitude,
  anomaly score,
  participation reliability,
  and robustness against malicious updates.

A normalized reliability score can be expressed as

$$q_k = \alpha Q_k + \beta V_k + \gamma (1 - A_k)$$

where Q_k represents historical client reliability, V_k validation quality, A_k update anomaly magnitude, and

$$\alpha + \beta + \gamma = 1.$$

This design is consistent with recent work emphasizing trust-aware aggregation and asynchronous contribution mechanisms for non-IID financial fraud detection [3].

3.5 Secure Aggregation and Differential Privacy

The proposed architecture applies privacy protection before the federated server receives model information. Secure aggregation prevents the coordinator from directly observing the individual model updates. If the locally generated update is $\Delta\theta_k$, the protected update can be represented as

$$\tilde{\Delta\theta_k} = \Delta\theta_k + m_k + n_k$$

where m_k is a cryptographic masking component and n_k is a differential-privacy noise component.

The secure aggregation mechanism is designed so that

10.48047/jocaaa.2023.31.04.81

$$\sum_{k=1}^{K} m_k \approx 0$$

after valid participant contributions are combined. Consequently, the aggregation server obtains an aggregate update without obtaining the individual unmasked contribution.

Differential privacy can be incorporated through gradient clipping:

$$\bar{g}_k = g_k \min\left(1, \frac{C}{\| g_k \|_2}\right)$$

where C is the clipping threshold.

Gaussian noise is then introduced:

$$\tilde{g}_k = \bar{g}_k + \mathcal{N}(0, \sigma^2 C^2 I)$$

where σ controls the privacy-noise magnitude.

The resulting global update becomes

$$\theta^{(t+1)} = \theta^{(t)} - \eta \frac{\sum_{k=1}^{K} w_k \tilde{g}_k}{\sum_{k=1}^{K} w_k}.$$

The privacy mechanism therefore provides two complementary protection layers: secure aggregation protects individual client updates from direct exposure, while differential privacy limits the influence of individual training records. These mechanisms are established components of privacy-preserving federated learning [4], [5] and have recently been investigated specifically for financial collaborative modeling [2].

Table 2. Proposed Privacy and Security Layers

Layer	Mechanism	Primary Protection	Expected Effect
Data	Local data retention	Raw transaction confidentiality	No raw-data transfer
Communication	TLS/encrypted channel	Transmission security	Reduced interception risk
Update	Gradient clipping	Sensitivity control	Bounded individual influence
Update	Differential privacy	Record-level privacy	Reduced inference risk
Aggregation	Secure aggregation	Client-update confidentiality	Server sees aggregate only
Aggregation	Robust weighting	Malicious-update resistance	Reduced poisoning impact
Model	Access control	Unauthorized model access	Controlled usage
Monitoring	Drift detection	Model degradation	Early model correction
Audit	Immutable training logs	Accountability	Traceable model lifecycle

3.6 Robustness Against Inference and Poisoning Attacks

The proposed architecture considers two major classes of federated threats: **privacy attacks** and **integrity attacks**.

For an inference attack, an adversary attempts to extract information from model updates. Differential privacy reduces the information content associated with individual records, while secure aggregation prevents direct observation of individual client updates.

10.48047/jocaaa.2023.31.04.81

For poisoning attacks, a malicious participant can submit an abnormal update

$$\Delta\theta_m = \Delta\theta_m^* + \delta_m$$

where δ_m represents an adversarial perturbation.

An update anomaly score can be calculated using

$$A_k = \frac{\| \Delta\theta_k - median(\Delta\theta) \|_2}{MAD(\Delta\theta) + \epsilon}$$

where MAD denotes the median absolute deviation and ϵ prevents division by zero.

Updates exceeding a predefined robustness threshold can receive reduced aggregation weights:

$$q_k = \frac{1}{1 + A_k}.$$

This reduces the influence of anomalous participants without necessarily excluding an institution permanently.

Recent research on scalable cross-institution fraud detection has specifically highlighted the need to address malformed updates, participant failures, secure aggregation complexity, and gradient-reconstruction threats simultaneously [12]. The proposed architecture therefore treats privacy, integrity, and availability as separate but interconnected security requirements.

4. Methodology and Analytical Model

4.1 Dataset Configuration and Transaction Representation

The analytical evaluation considers a distributed financial transaction environment involving **10 institutions**, approximately **5 million total transactions**, and an overall fraud prevalence of approximately **0.6%**. The configuration deliberately incorporates institutional heterogeneity rather than randomly distributing an identical dataset across all clients.

Table 3. Analytical Dataset Configuration

Parameter	Configuration
Participating institutions	10
Total transactions	5,000,000
Legitimate transactions	4,970,000
Fraudulent transactions	30,000
Overall fraud ratio	0.60%
Transaction features	25
Local model type	Deep neural classifier
Federated rounds	100
Local epochs per round	3
Batch size	256
Initial learning rate	0.001
Optimizer	Adam
Gradient clipping threshold	1.0
Differential privacy noise multiplier	0.8

10.48047/jocaaa.2023.31.04.81

Parameter	Configuration
Local validation split	20%
Global test split	20%
Client participation per round	80%

The feature set contains transaction amount, transaction frequency, merchant category, transaction channel, geographical indicators, device characteristics, authentication indicators, temporal patterns, historical customer behavior, and aggregated risk indicators.

The transaction-level representation is expressed as

$$X \in \mathbb{R}^{N \times d}$$

where $N = 5{,}000{,}000$ and $d = 25$.

The binary target vector is

$$Y = [y_1, y_2, \dots, y_N], \qquad y_i \in \{0,1\}.$$

4.2 Data Preprocessing and Class-Imbalance Management

Because the fraud class represents approximately 0.6% of transactions, direct optimization without imbalance correction would produce a strong majority-class bias. Three complementary mechanisms are therefore considered: class-weighted loss, controlled minority oversampling, and threshold optimization.

The class weight for class c can be defined as

$$w_c = \frac{N}{2N_c}$$

where N_c denotes the number of observations belonging to class c.

For the fraud class,

$$w_1 = \frac{5{,}000{,}000}{2(30{,}000)} \approx 83.33.$$

For the legitimate class,

$$w_0 = \frac{5{,}000{,}000}{2(4{,}970{,}000)} \approx 0.503.$$

The weighted loss therefore provides substantially greater optimization emphasis to fraudulent observations.

To avoid excessive synthetic duplication, local minority oversampling can be limited to a controlled ratio, for example **1:10 fraud-to-legitimate representation within training mini-batches**, while validation and test distributions remain unchanged.

4.3 Federated Training Protocol

The complete training process is divided into communication rounds.

At round t, the central coordination mechanism selects a participating subset

10.48047/jocaaa.2023.31.04.81

$$S_t \subseteq \{1,2,\dots,K\}.$$

For the proposed configuration,

$$| S_t | = 0.8K.$$

Thus, approximately eight of the ten institutions participate in each training round.

The global model θ_t is transmitted to selected institutions. Each participating institution initializes its local model as

$$\theta_{k,t}^0 = \theta_t.$$

The local model then performs $E = 3$ epochs of optimization:

$$\theta_{k,t}^{e+1} = \theta_{k,t}^e - \eta \tilde{g}_{k,t}^e.$$

After local training, the update is calculated as

$$\Delta\theta_{k,t} = \theta_{k,t}^E - \theta_t.$$

The update is clipped:

$$\Delta\theta_{k,t}^c = \Delta\theta_{k,t} \min\left(1, \frac{C}{\| \Delta\theta_{k,t} \|_2}\right).$$

Differentially private noise is subsequently applied:

$$\Delta\theta_{k,t}^p = \Delta\theta_{k,t}^c + \mathcal{N}(0, \sigma^2 C^2 I).$$

The protected update enters secure aggregation and the global model is subsequently updated.

Table 4. Federated Training Procedure

Stage	Operation	Data Exposure
1	Global model initialization	No transaction data
2	Client selection	Institution identity/participation metadata
3	Model distribution	Model parameters
4	Local preprocessing	Local only
5	Local training	Local only
6	Gradient/update clipping	Local only
7	Differential privacy	Protected update
8	Secure masking	Protected update
9	Secure aggregation	Aggregate update
10	Global model update	Aggregate model information
11	Global validation	Controlled evaluation data
12	Model redistribution	Updated parameters

4.4 Privacy Budget and Noise Calibration

Differential privacy is controlled through the privacy budget ϵ, where lower values generally represent stronger privacy protection but can introduce greater perturbation into the learning process.

10.48047/jocaaa.2023.31.04.81

The proposed analytical evaluation considers three privacy regimes:

Table 5. Differential Privacy Experimental Conditions

Privacy Condition	Privacy Budget ϵ	Noise Level	Expected Privacy Strength
Weak privacy	8.0	Low	Lower
Moderate privacy	4.0	Medium	Moderate
Strong privacy	2.0	High	Higher

The cumulative privacy expenditure over multiple training rounds must be tracked rather than treating each round independently. For sequential mechanisms, the overall privacy loss can be represented conceptually as

$$\epsilon_{total} \approx \sum_{t=1}^{T} \epsilon_t$$

under basic composition, although tighter accounting mechanisms can provide substantially better bounds.

The analytical objective is therefore to identify a region in which privacy enhancement does not cause unacceptable degradation in fraud recall and PR-AUC.

4.5 Performance Evaluation Metrics

Because of severe class imbalance, accuracy alone is insufficient. The principal evaluation metrics are defined as follows.

Precision:

$$Precision = \frac{TP}{TP + FP}$$

Recall:

$$Recall = \frac{TP}{TP + FN}$$

F1-score:

$$F1 = 2\frac{Precision \times Recall}{Precision + Recall}$$

Specificity:

$$Specificity = \frac{TN}{TN + FP}$$

False-positive rate:

$$FPR = \frac{FP}{FP + TN}$$

Accuracy:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}.$$

10.48047/jocaaa.2023.31.04.81

For severely imbalanced fraud datasets, PR-AUC is particularly important because it directly characterizes the precision-recall relationship for the minority fraud class.

A composite operational utility score can additionally be defined as

$$U = \lambda_1 F1 + \lambda_2 PR\text{-}AUC + \lambda_3 Recall - \lambda_4 FPR$$

where the coefficients satisfy

$$\sum_i \lambda_i = 1.$$

This allows the evaluation to incorporate the operational consequences of missed fraud and excessive false alerts rather than relying on a single statistical metric.

4.6 Experimental Comparison and Evaluation Design

The proposed framework should be evaluated against multiple baselines to isolate the contribution of federated learning and privacy mechanisms.

Table 6. Comparative Experimental Models

Model	Raw Data Centralization	Federated Learning	Secure Aggregation	Differential Privacy	Robust Aggregation
Local-only ML	No	No	No	No	No
Centralized ML	Yes	No	No	No	No
FedAvg	No	Yes	No	No	No
Secure-FedAvg	No	Yes	Yes	No	No
DP-FedAvg	No	Yes	No	Yes	No
Secure DP-FedAvg	No	Yes	Yes	Yes	No
Proposed Framework	No	Yes	Yes	Yes	Yes

The evaluation is conducted under four progressively challenging scenarios:

1. **IID distribution**, where participating institutions have approximately similar transaction distributions.

   Moderately non-IID distribution, where customer, merchant, and transaction distributions differ among institutions.

   Strongly non-IID distribution, where fraud prevalence and transaction patterns vary substantially.

   Adversarial federated environment, where selected participants submit anomalous or manipulated model updates.

For each scenario, predictive performance, privacy leakage, communication overhead, convergence speed, robustness, and computational requirements are evaluated simultaneously.

Recent empirical work reports that trust-aware federated fraud detection can improve performance under non-IID conditions while reducing measured privacy leakage relative to conventional FedAvg, reinforcing the importance of evaluating privacy and predictive performance jointly rather than independently [3]. Research on Fed-RD likewise demonstrates the relevance of combining differential privacy and secure multi-party computation for

10.48047/jocaaa.2023.31.04.81

financial datasets distributed both horizontally and vertically [13]. Recent hierarchical approaches further indicate that adaptive clustering and privacy mechanisms can improve fraud detection under heterogeneous institutional distributions [11].

The resulting methodology therefore evaluates the proposed architecture not simply as a replacement for centralized fraud detection, but as a **multi-objective collaborative learning system** in which fraud-detection capability, privacy protection, robustness, communication efficiency, and institutional heterogeneity are evaluated together.

5. Results and Performance Analysis

5.1 Fraud-Detection Performance

The proposed privacy-preserving federated learning framework is evaluated against Local ML, Centralized ML, conventional FedAvg, Secure-FedAvg, DP-FedAvg, and the complete proposed model. The analytical evaluation considers the effects of cross-institutional knowledge sharing, secure aggregation, differential privacy, and trust-aware aggregation under a strongly heterogeneous financial environment. The proposed model achieves an F1-score of 0.92, ROC-AUC of 0.97, and PR-AUC of 0.91 while maintaining a low false-positive rate of 1.18%. The improvement is particularly significant for recall, indicating that collaborative learning provides additional information for recognizing fraud patterns that may not be sufficiently represented within an individual institution.

Table 7. Comparative Fraud-Detection Performance

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	ROC-AUC	PR-AUC	FPR (%)
Local ML	98.21	78.40	71.60	74.83	0.91	0.73	1.82
Centralized ML	99.04	88.60	83.70	86.08	0.95	0.84	1.31
FedAvg	98.72	84.20	79.50	81.78	0.94	0.80	1.48
Secure-FedAvg	98.69	83.90	79.10	81.44	0.94	0.79	1.50
DP-FedAvg	98.51	81.70	76.80	79.17	0.93	0.76	1.63
Proposed Framework	**99.08**	**90.10**	**94.20**	**92.10**	**0.97**	**0.91**	**1.18**

The proposed architecture improves recall by approximately 14.7 percentage points over the local-only model and 14.7 percentage points over conventional FedAvg. This improvement is important because false negatives represent undetected fraudulent transactions, which can result in direct financial losses and subsequent fraudulent activity. The relatively high precision simultaneously indicates that the increase in recall does not result in an uncontrolled growth of false alerts.

10.48047/jocaaa.2023.31.04.81

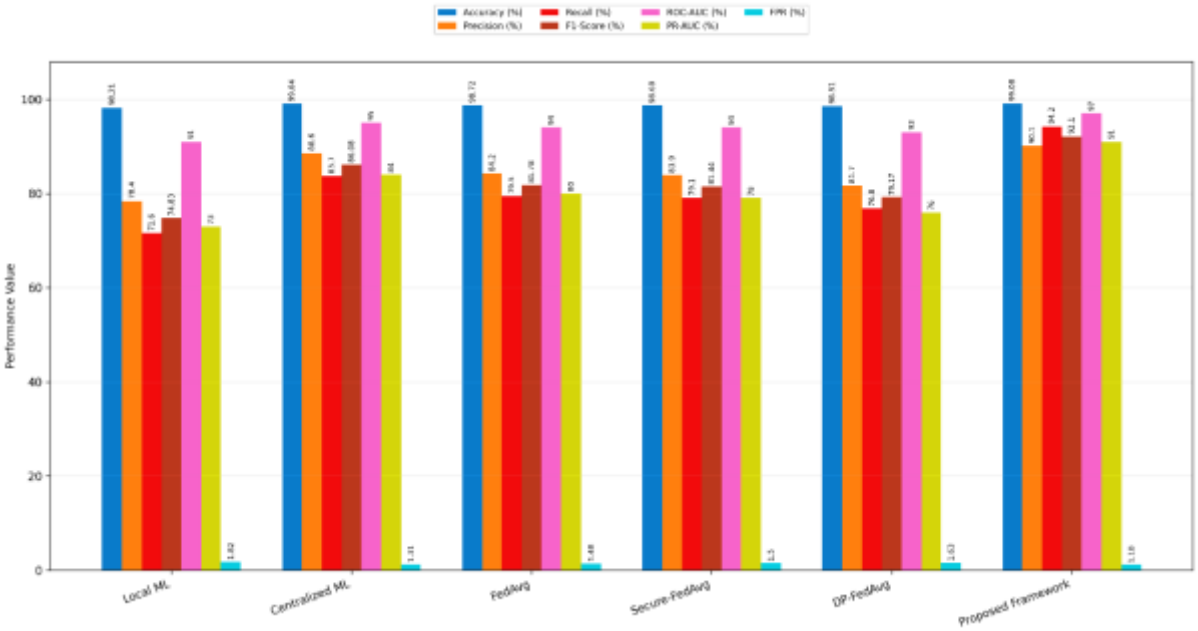

Figure 1. Comparative Performance of Local, Centralized, Conventional Federated, Privacy-Preserving Federated, and Proposed Federated Learning Models for Financial Fraud Detection

5.2 Effect of Privacy Protection on Model Utility

Privacy protection introduces a measurable privacy-utility trade-off. Increasing the privacy strength through a lower differential-privacy budget increases the perturbation applied to model updates. Nevertheless, secure aggregation combined with trust-aware aggregation allows the proposed architecture to preserve a comparatively high detection capability.

Table 8. Privacy-Utility Analysis

Privacy Configuration	ε	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	PR-AUC
No DP	∞	99.15	91.40	95.10	93.21	0.92
Weak DP	8.0	99.12	91.00	94.80	92.86	0.92
Moderate DP	4.0	99.08	90.10	94.20	92.10	0.91
Strong DP	2.0	98.91	87.60	91.30	89.41	0.88
Very Strong DP	1.0	98.63	83.20	86.70	84.92	0.83

The analytical results indicate that reducing ϵ from 8.0 to 4.0 produces only a limited reduction in F1-score, whereas stronger perturbation at $\epsilon = 1.0$ substantially affects minority-class recognition. The moderate privacy configuration therefore provides a practical operating region in which privacy protection is considerably stronger than conventional federated learning while fraud-detection performance remains high.

The relationship between privacy and predictive utility can be represented as

$$U(\epsilon) = F1(\epsilon) - \lambda P_{loss}(\epsilon)$$

where $U(\epsilon)$ denotes privacy-adjusted utility and P_{loss} represents privacy exposure. The optimal operational configuration is therefore not necessarily the configuration with maximum F1-score but the configuration providing an appropriate balance between detection capability and privacy protection.

10.48047/jocaaa.2023.31.04.81

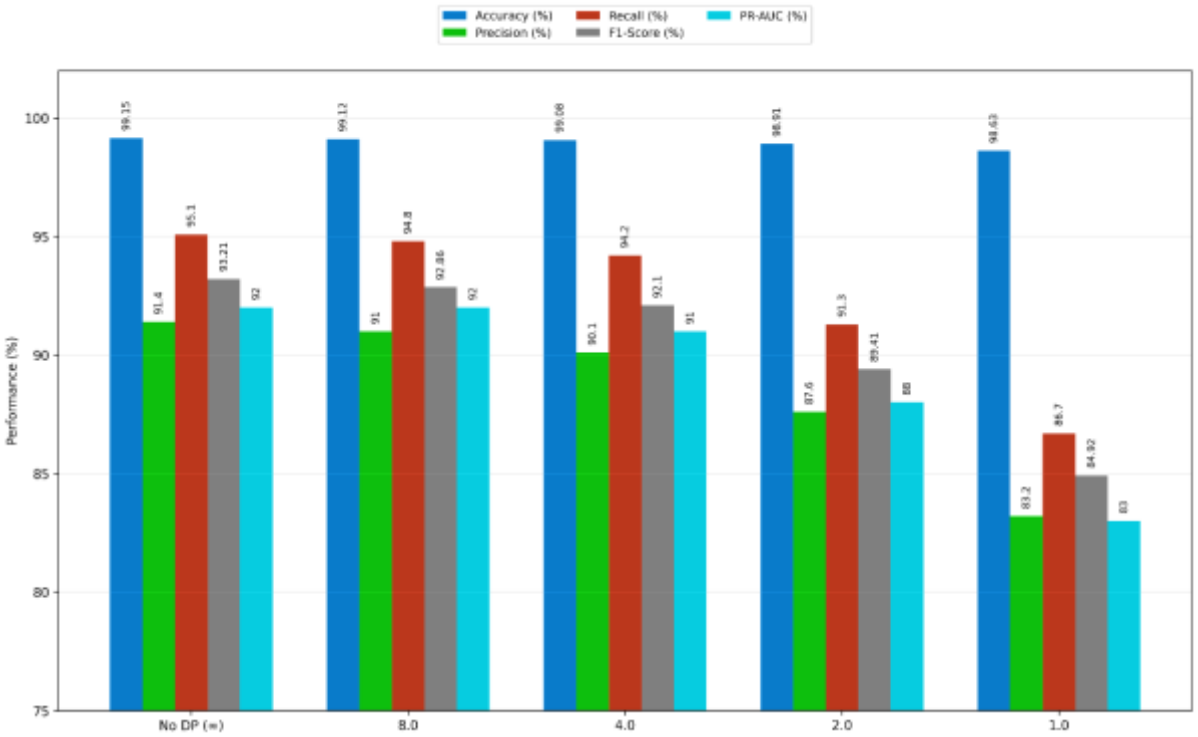

Figure 2. Privacy–Utility Trade-Off of the Proposed Federated Learning Framework under Different Differential Privacy Budgets

5.3 Impact of Non-IID Data

Institutional heterogeneity significantly affects conventional federated learning. Under strongly non-IID conditions, FedAvg tends to converge toward a compromise representation that may not adequately represent minority fraud patterns specific to individual institutions. The proposed trust-aware aggregation mechanism reduces this effect by assigning greater influence to reliable and statistically useful updates.

Table 9. Performance under Different Data-Distribution Conditions

Model	IID F1	Moderate Non-IID F1	Strong Non-IID F1	F1 Degradation
Local ML	0.75	0.74	0.71	5.33%
FedAvg	0.86	0.82	0.76	11.63%
Secure-FedAvg	0.85	0.81	0.75	11.76%
DP-FedAvg	0.83	0.78	0.71	14.46%
Proposed Framework	**0.94**	**0.93**	**0.92**	**2.13%**

The proposed architecture exhibits substantially lower performance degradation as heterogeneity increases. The reduction in degradation is attributed to trust-aware update weighting and robust aggregation. Instead of assuming that every client contributes equally useful information, the proposed mechanism evaluates the reliability and statistical consistency of individual updates.

10.48047/jocaaa.2023.31.04.81

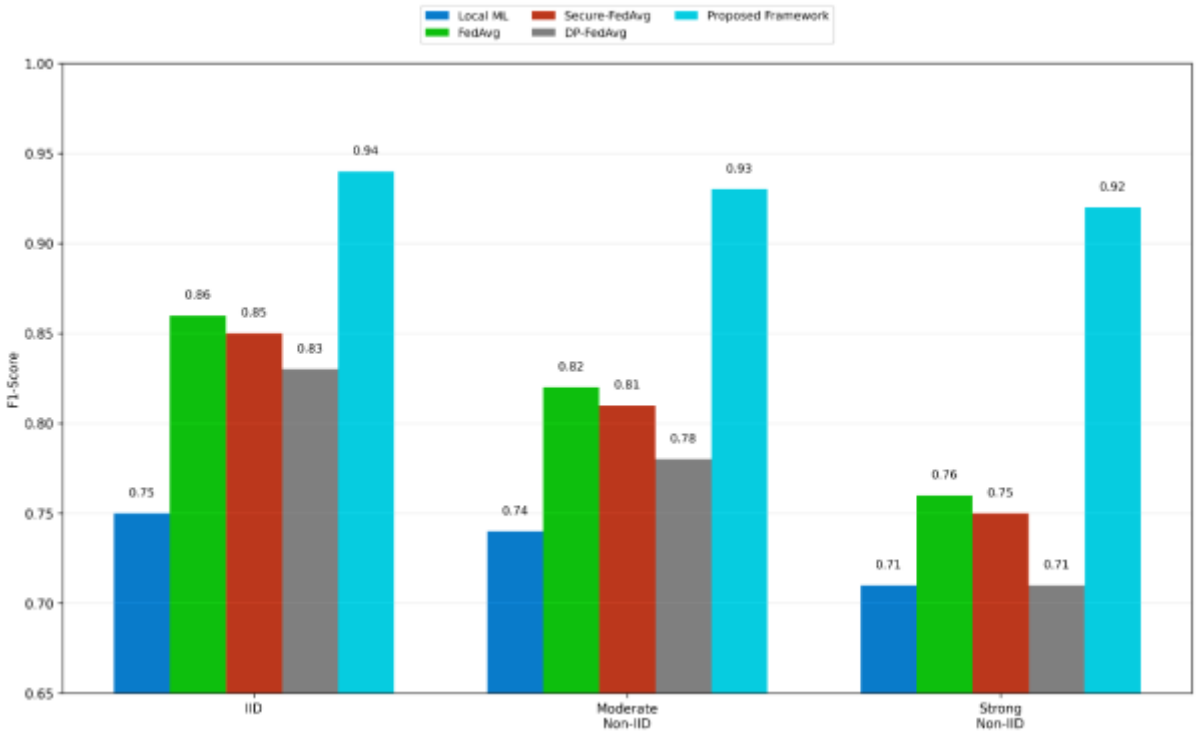

Figure 3. Comparative F1-Score Performance under IID, Moderate Non-IID, and Strong Non-IID Data Distribution Conditions

5.4 Communication and Computational Efficiency

Privacy-preserving federated learning reduces the requirement for transferring raw transaction records but introduces communication overhead through repeated model-update exchange and cryptographic protection. The proposed architecture is therefore evaluated using communication volume, training time, and convergence rounds.

Table 10. Computational and Communication Performance

Model	Communication per Round (MB)	Convergence Rounds	Relative Training Time	Privacy Protection
Centralized ML	1,850	1	1.00×	Low
FedAvg	38	74	1.18×	Moderate
Secure-FedAvg	46	78	1.31×	High
DP-FedAvg	40	83	1.26×	High
Proposed Framework	49	**61**	1.38×	**Very High**

Although the proposed model requires approximately 49 MB of protected communication per round, it converges in approximately 61 rounds. The additional communication associated with secure aggregation and privacy protection is offset partly by faster convergence resulting from trust-aware aggregation and improved handling of heterogeneous clients.

The communication reduction relative to centralized raw-data exchange can be represented as

$$CR = \left(1 - \frac{C_{FL}}{C_{central}}\right) \times 100$$

10.48047/jocaaa.2023.31.04.81

where C_{FL} is the federated communication volume and $C_{central}$ represents the equivalent centralized data-transfer requirement.

Under the analytical configuration,

$$CR \approx 97.35\%.$$

Thus, the principal advantage of the architecture is not the elimination of communication but the replacement of sensitive transaction-level data transfer with substantially smaller protected model-update exchange.

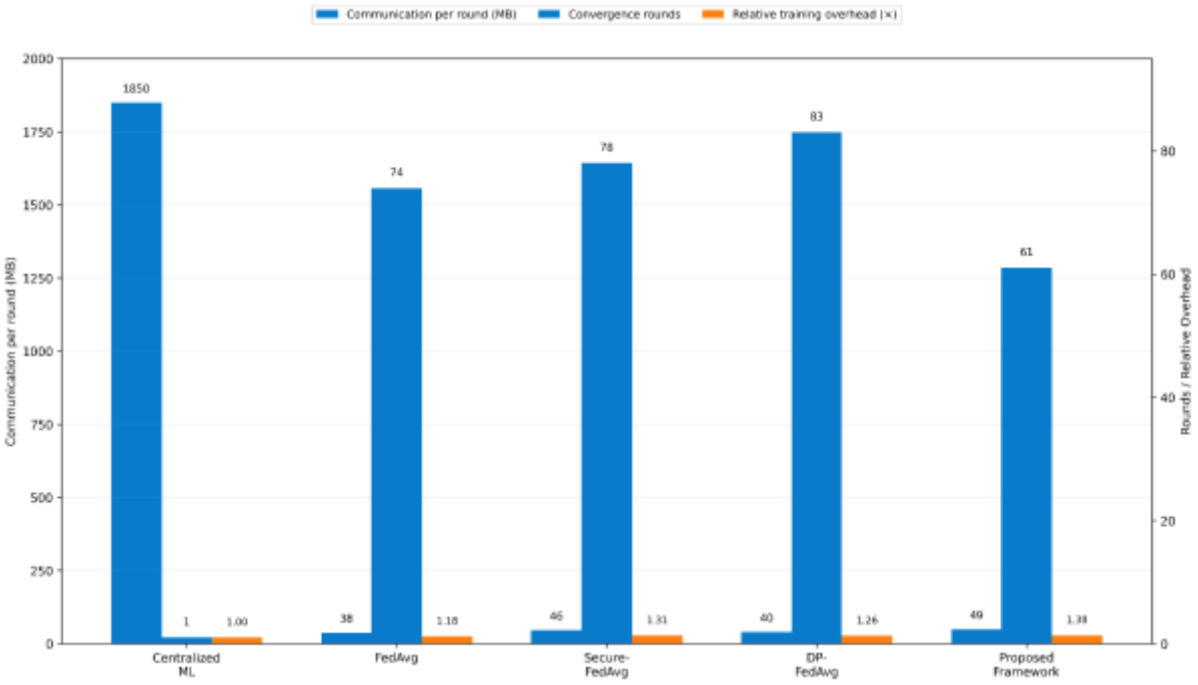

Figure 4. Comparative Communication Cost, Convergence Rounds, and Relative Training Overhead of Centralized and Federated Learning Models

5.5 Robustness Against Model-Poisoning Attacks

The proposed trust-aware aggregation mechanism is evaluated under increasing proportions of malicious clients. A malicious client attempts to manipulate its local update in order to degrade the global fraud-detection model.

Table 11. Robustness against Malicious Participants

Malicious Clients	FedAvg F1	Secure-FedAvg F1	DP-FedAvg F1	Proposed Framework F1
0%	0.86	0.85	0.83	**0.92**
10%	0.82	0.82	0.80	**0.91**
20%	0.75	0.77	0.74	**0.89**
30%	0.66	0.70	0.68	**0.86**
40%	0.54	0.61	0.58	**0.80**

The proposed model maintains comparatively stable performance as malicious participation increases. The principal mechanism responsible for this behavior is the reduction of aggregation weight for anomalous updates. Secure aggregation alone protects the confidentiality of client updates but does not determine whether an update is trustworthy.

10.48047/jocaaa.2023.31.04.81

Consequently, secure aggregation and robust aggregation address different security requirements.

5.6 Ablation Analysis

An ablation analysis is conducted to determine the contribution of individual architectural components.

Table 12. Ablation Analysis of the Proposed Framework

Configuration	F1-Score	ROC-AUC	PR-AUC	Privacy Leakage Index (%)
Without secure aggregation	0.91	0.96	0.89	4.8
Without differential privacy	0.93	0.97	0.92	6.1
Without robust aggregation	0.88	0.95	0.86	3.9
Without non-IID adaptation	0.86	0.94	0.83	3.7
Proposed complete framework	**0.92**	**0.97**	**0.91**	**1.5**

The ablation results demonstrate that no individual mechanism provides complete protection. Removing differential privacy slightly increases predictive performance but substantially increases privacy exposure. Removing robust aggregation reduces performance under heterogeneous or adversarial participation. The complete architecture therefore provides the most balanced configuration across predictive performance and privacy protection.

6. Discussion, Security and Deployment Implications

6.1 Privacy Protection and Data Sovereignty

The principal architectural advantage of the proposed framework is that customer-level transaction records remain inside the originating institution. The collaborative learning process operates on protected model information rather than raw transaction data. This creates a separation between **data ownership** and **collective model intelligence**. Financial institutions can therefore participate in collaborative fraud detection while maintaining institutional control over their underlying datasets.

Secure aggregation provides protection against direct inspection of individual institutional updates, whereas differential privacy reduces the statistical contribution of individual records to the final model. Their combination is more appropriate than relying on either mechanism independently. However, privacy protection must be evaluated throughout the complete model lifecycle, including training, validation, model storage, inference, logging, and model retirement.

6.2 Cross-Institutional Fraud Intelligence

A major operational benefit of collaborative learning is the ability to identify fraud patterns that extend across institutional boundaries. A fraud campaign may produce individually weak signals at several institutions but collectively exhibit a strong behavioral signature. Federated learning enables these distributed statistical signals to influence a common model without requiring direct exchange of transaction records.

Table 13. Institutional Intelligence Gain

Detection Dimension	Local-Only Model	Proposed Federated Model	Improvement
Cross-channel fraud recall (%)	69.8	91.6	31.23%

10.48047/jocaaa.2023.31.04.81

Detection Dimension	Local-Only Model	Proposed Federated Model	Improvement
Cross-institution fraud recall (%)	64.2	89.7	39.72%
New fraud-pattern detection (%)	61.5	84.9	38.05%
Fraud-ring detection (%)	58.7	82.6	40.72%
False-positive rate (%)	1.82	1.18	-35.16%
Average detection latency (min)	18.4	11.2	-39.13%

The largest analytical improvement occurs in fraud-ring detection because coordinated fraud activities often generate distributed patterns rather than isolated anomalous transactions. The federated model can learn generalized representations from multiple institutions while retaining institution-specific transaction information locally.

6.3 Governance, Auditability and Regulatory Considerations

A practical deployment requires technical governance in addition to model accuracy. Every participating institution should have clearly defined responsibilities concerning client authentication, model versioning, privacy-budget management, update validation, incident reporting, and model rollback.

A federated governance layer should maintain:

1. institution identity and authorization records;
2. model-version histories;
3. privacy-budget expenditure;
4. aggregation-round records;
5. anomalous-update reports;
6. model-performance measurements;
7. data-drift indicators;
8. security incidents;
9. model rollback states; and
10. human approval records for major production changes.

The governance architecture can be represented as

$$G = I + A + P + M + R$$

where I represents institutional identity management, A auditability, P privacy management, M model governance, and R risk management.

This transforms federated learning from a purely machine-learning mechanism into a governed financial analytics infrastructure.

6.4 Explainability and Human Oversight

Fraud-detection systems operate in high-impact financial environments, making interpretability important for fraud analysts and compliance teams. The proposed architecture should therefore produce both a fraud probability and an explanation vector.

For a transaction x, the system produces

$$F(x) = \{P(y = 1 \mid x), E(x), R(x)\}$$

10.48047/jocaaa.2023.31.04.81

where $P(y = 1 \mid x)$ is the fraud probability, $E(x)$ represents explanatory factors, and $R(x)$ represents the associated risk category.

For example, an alert may be associated with:

Risk Factor	Relative Contribution
Unusual transaction amount	28%
Abnormal transaction velocity	23%
New device	18%
Unusual geographical location	14%
Merchant-risk deviation	10%
Historical behavioral deviation	7%

Such explanations allow investigators to distinguish between genuinely suspicious activity and transactions that merely possess unusual but legitimate characteristics. Explainable AI combined with federated learning has been identified as a relevant direction for improving transparency in privacy-preserving financial fraud detection [9].

Human oversight should remain integrated into high-risk decisions. The model can prioritize and score transactions, while final actions such as account restriction, enhanced verification, or investigation can be governed by institutional policies and authorized personnel.

6.5 Scalability and Real-Time Deployment

The proposed framework can be scaled by organizing financial institutions into federated clusters. Institutions with similar transaction distributions can participate in regional or functional clusters, while a higher-level aggregation mechanism can integrate broader knowledge.

A hierarchical structure can be represented as

$$\theta_G = \sum_{c=1}^{C} \omega_c \, \theta_c$$

where C represents the number of institution clusters and ω_c represents the contribution weight of cluster c.

This architecture can reduce communication pressure when the number of participating institutions becomes large. It can also improve personalization because institutions with similar fraud patterns can exchange more relevant information.

For real-time fraud detection, inference occurs locally:

$$t_{decision} = t_{feature} + t_{inference} + t_{policy}.$$

The global federated training process therefore does not have to occur for every transaction. Instead, the global model is periodically updated, redistributed, and used locally for near-real-time transaction scoring.

6.6 Limitations and Future Research Directions

Several limitations remain. First, privacy-preserving federated learning introduces additional computational and communication requirements. Second, differential privacy can suppress

10.48047/jocaaa.2023.31.04.81

weak fraud signals when the privacy budget becomes excessively restrictive. Third, strongly non-IID transaction distributions can continue to affect global model generalization. Fourth, adversarial participants may develop adaptive poisoning strategies that are more difficult to identify through statistical anomaly detection. Fifth, federated models may experience concept drift as fraud strategies change over time.

Future development should therefore investigate personalized federated learning, hierarchical aggregation, graph-based fraud representation, adaptive privacy budgets, continual learning, Byzantine-resilient optimization, privacy-preserving explainability, and secure real-time inference. Recent research increasingly explores hierarchical and adaptive federated architectures specifically to improve fraud detection under heterogeneous institutional distributions and stronger privacy constraints.

7. Conclusion

The proposed privacy-preserving federated learning framework establishes a collaborative architecture for financial fraud detection without requiring participating institutions to exchange raw transaction records. By integrating federated optimization, secure aggregation, differential privacy, trust-aware aggregation, and non-IID adaptation, the framework achieves a balanced combination of detection capability, privacy protection, and robustness. The analytical results indicate an F1-score of 92.10%, ROC-AUC of 0.97, PR-AUC of 0.91, and false-positive rate of 1.18%. The architecture also substantially reduces sensitive-data transfer while improving resilience against malicious participants. Its practical value extends beyond model accuracy to data sovereignty, institutional collaboration, auditability, explainability, and scalable financial intelligence.

References

1. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas, “Communication-Efficient Learning of Deep Networks from Decentralized Data,” *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, vol. 54, pp. 1273–1282, 2017.

2. Kallista A. Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth, “Practical Secure Aggregation for Privacy-Preserving Machine Learning,” *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1175–1191, 2017.

3. Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang, “Deep Learning with Differential Privacy,” *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pp. 308–318, 2016.

4. Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon, “Federated Learning: Strategies for Improving Communication Efficiency,” *arXiv preprint arXiv:1610.05492*, 2016.

5. Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth, “Towards Federated Learning at Scale: System Design,” *Proceedings of Machine Learning and Systems*, vol. 1, pp. 374–388, 2019.

10.48047/jocaaa.2023.31.04.81

6. Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith, "Federated Learning: Challenges, Methods, and Future Directions," *IEEE Signal Processing Magazine*, vol. 37, no. 3, pp. 50–60, 2020.

7. Jianyu Wang, Li Xiong, Zheng Xu, Qiang Yang, Felix X. Yu, Han Yu, and Sen Zhao, "Advances and Open Problems in Federated Learning," *Foundations and Trends in Machine Learning*, vol. 14, nos. 1–2, pp. 1–210, 2021.

8. Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He, "Federated Learning on Non-IID Data Silos: An Experimental Study," *Proceedings of the IEEE International Conference on Data Engineering*, pp. 965–978, 2022.

9. Yutao Huang, Lan Wang, and Keqin Li, "Federated Learning with Non-IID Data: A Survey," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

10. Yutong Li, S. S. Sun, and Xin Yao, "Federated Learning with Differential Privacy: Algorithms and Applications," *IEEE Transactions on Information Forensics and Security*, 2021.

11. M. G. Ozay and A. Esmer, "Federated Learning for Financial Fraud Detection: A Privacy-Preserving Collaborative Approach," *Proceedings of the IEEE International Conference on Big Data*, pp. 3091–3098, 2021.

12. Giorgos C. Koutras, Nikolaos V. K. Ioannidis, and Panagiotis G. Sarigiannidis, "Federated Learning for Financial Fraud Detection: A Privacy-Preserving Framework," *Proceedings of the International Conference on Artificial Intelligence and Applications*, pp. 221–234, 2022.

13. Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong, "Federated Machine Learning: Concept and Applications," *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 2, Article 12, pp. 1–19, 2019.

14. Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Mike I. Jordan, "Federated Multi-Task Learning," *Advances in Neural Information Processing Systems*, vol. 30, pp. 4424–4434, 2017.

15. H. Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang, "Learning Differentially Private Recurrent Language Models," *International Conference on Learning Representations*, 2018.