IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

A Scalable Federated Learning Architecture for Privacy-Preserving Financial Data Processing

Praveen Kumar Reddy Gouni[1], Mohammed Abdul Faheem[2]

Department of Information Technology, Belhaven University, MS, USA[1]

Department of Information Technology, Trine University, IN, USA[2]

Abstract: Due to the quick digital transformation of the banking and financial services sector, financial data is now larger, more sensitive, and easier to find. Conventional centralized machine learning solutions pose serious privacy, security, and regulatory issues to businesses since they need gathering user data in one place. Federated Learning (FL) is a novel method that enables users to train models collectively without exchanging raw data. The PCI-DSS, GDPR, and RBI privacy laws are all met by this. In addition to incorporating features like safe aggregation, homomorphic encryption, differential privacy, and blockchain-based auditability, this paper provides a full federated learning system for financial analytics that protects privacy. The study illustrates how FL might reduce the risks associated with central data storage, hence enhancing financial forecasts, risk modelling, fraud detection, and credit rating. A thorough experimental analysis is presented to compare FL to conventional centralized approaches on important performance metrics as computation load, accuracy, privacy protection, and communication efficiency. The results demonstrate that FL maintains model performance competitiveness while significantly improving privacy and regulatory compliance. Additionally, in distributed financial contexts, the suggested blockchain-based auditability layer guarantees the permanence of transparency, verifiability, and recording. The essay also covers potential difficulties, how to apply the concepts, and the most effective methods for handling actual financial systems. Our study concludes by demonstrating the great potential of Federated Learning as a safe, scalable, and legally permissible alternative for next-generation financial analytics.

Keywords: blockchain auditability, risk modelling, homomorphic encryption, secure aggregation, distributed machine learning, federated learning, credit scoring, financial fraud detection, and privacy-preserving analytics.

I. INTRODUCTION

Online banking, fintech platforms, mobile payments, and AI-powered decision systems are all becoming more and more common, and this is causing a significant digital revolution in the financial sector [1]. In order to detect fraud, check credit, categorize consumers, report to regulators, and evaluate risk, banks and other financial institutions are increasingly utilizing data-driven models. As a result, the amount and sensitivity of financial data have significantly expanded. Predictive model training usually involves gathering raw data from several sources and combining it in one place. Although productivity is increased by this centralized approach, security is jeopardized since it raises the possibility of data breaches, unauthorized access, insider threats, and a disrespect for stringent privacy rules.

It is very challenging to exchange, store, and process financial records across borders due to laws like the General Data Protection Regulation (GDPR), India's Digital Personal Data Protection (DPDP) Act, the California Consumer Privacy Act (CCPA), and PCI-DSS. Furthermore, banks are less likely to work together to exchange data because they are in competition with one another [2]. As a result, machine learning becomes less useful and datasets are less complete. These problems show that protecting consumer data and lawful corporate practices are highly at odds. One novel method for training machine learning models at several locations without sending raw data is federated learning (FL). Only modifications to the model's parameters or gradients are transmitted. In addition to maintaining data in its current location, this lessens privacy problems

The security, privacy, and accountability standards that contemporary financial services need are met by FL. Fintech firms and banks may work together on predictive algorithms while keeping ownership over their own datasets thanks to FL [3]. This opens up new avenues to enhance AML monitoring, enhance fraud detection, decrease bias in credit scoring, and improve risk management.

This article explores a comprehensive, privacy-preserving federated learning architecture that includes homomorphic encryption, safe aggregation, differential privacy, and blockchain-based auditability in order to satisfy the growing demands of the financial analytics industry [4].

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

II. PROBLEM STATEMENT

In the financial services industry, machine learning is playing a bigger role in spotting fraud, assessing creditworthiness, spotting dangers, and making sure regulations are followed [5]. Nevertheless, substantial, varied, and representative datasets are needed to create quality models. Few of these assets are owned by individual banks, fintech firms, or payment processors. For organizations looking to work together on analytics and successfully use AI, this poses serious challenges.

A. Data Privacy Legislation

The CCPA, GDPR, DPDP Act, and PCI-DSS are just a few of the strict regulations that banks and other financial institutions must follow to protect personal data [6]. These rules outline how you will handle, store, and distribute customer data. Consequently, it is not practical to build centralized databases for collaborative model training since entities are not allowed to exchange data directly.

B. Risks to Security in Centralized ML

Although centralized machine learning systems have advantages, they create data silos that are vulnerable to hacking, insider threats, illegal access, and significant data breaches. Reputation and operations suffer when a single point of failure jeopardizes millions of confidential financial records [7].

C. Distinct and interconnected data silos

Financial institutions, including banks, employ a variety of formats, schemas, storage systems, and quality requirements [8]. This fragmentation reduces model performance, skews credit scores due to incomplete histories, and makes it more challenging to identify patterns of fraud across institutions.

D. Challenges for Organization and Competition

Because of worries about intellectual property, competition, and the disclosure of important corporate information, institutions are reluctant to provide private financial records [9]. AI's ability to advance concurrently is hampered by these structural problems.

E. The fundamental issue statement

How can banks and other financial organizations work together to develop machine learning models that are safe, dependable, and compliant while preserving private and sensitive data [10].

Table 1: Key Challenges in Traditional Financial Machine Learning

Challenge Category	Description	Impact on ML Development
Privacy & Compliance	Legal restrictions on data sharing	Prevents cross-institution training
Security Risks	Centralized storage vulnerabilities	Increases breach and attack surface
Data Fragmentation	Non-uniform, isolated datasets	Leads to biased or weak models
Competitive Barriers	Reluctance to share proprietary data	Limits collaboration and innovation
Infrastructure Diversity	Varied systems across institutions	Reduces interoperability and scalability

III. RESEARCH GAPS

Federated Learning (FL) is still relatively new and underutilized in the banking sector, despite the tremendous expansion of machine learning applications [11]. Although earlier studies have produced promising results, there are still a lot of unanswered problems before it can be expanded, put into practice in the real world, and guaranteed to be compliant with the law. The following are the main areas of research that needed to be looked into.

A. The number of FL designs with a focus on finance is somewhat lacking

For financial tasks like credit risk modelling, derivative pricing, fraud detection, and anti-money laundering, the majority of FL frameworks are not designed to be employed. Non-standard financial data architectures, high-frequency transaction streams, and unbalanced fraud datasets cannot be handled by current systems [12]. This difference necessitates model designs that are tailored to financial analytics and based on domain knowledge.

B. Inadequate application of contemporary privacy safeguards

Reconstruction attacks may still be able to obtain important information even if FL protects the original material. A cohesive financial-grade architecture incorporating several potent privacy-preserving methods, including safe aggregation,

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

homomorphic encryption (HE), differential privacy (DP), and secure multiparty computation (SMC), was rare in earlier research [13]. It is less useful in practice because it does not guarantee the privacy of all individuals.

C. Limited assessment based on non-IID and data from multiple institutions

The financial information of banks varies substantially based on the kinds of customers they serve, their business practices, their risk management guidelines, and the configuration of their systems. The majority of FL studies either employ basic datasets or make the assumption that everything is IID. Accordingly, the models are not applicable to actual financial circumstances. A large number of thorough studies on realistic, non-IID multi-bank datasets are lacking [14].

D. Insufficient regulatory auditability and traceability

AI systems that are easy to comprehend, use, and audit are preferred by financial authorities. To comply with regulations like the GDPR, DPDP, and Basel III risk recommendations, a large portion of the FL research now in progress does not employ blockchain or tamper-proof audit trails [15]. People are reluctant to embrace it since it cannot be audited in a way that complies with regulations.

E. Inadequate analysis of the computation and transmission costs

Communication between clients and servers is made more challenging by FL's encryption and privacy capabilities [16]. Additionally, it increases the demand on computing and complicates communication. Because most studies do not contain these trade-offs in financial-scale databases, institutions find it difficult to profit from them.

The significance of a thorough FL framework for financial analytics that complies with legal requirements and safeguards individuals' privacy is highlighted by these research gaps.

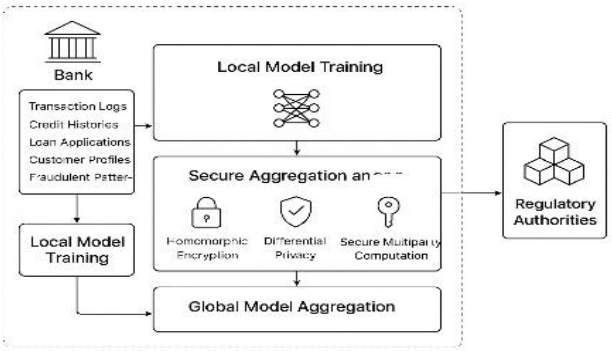

Figure 1: Federated Learning Architectures

IV. LITERATURE REVIEW

Over the past decade, federated learning (FL) and artificial intelligence (AI) have gained popularity in the field of financial analytics [17]. The research is still dispersed, with many papers focusing only on fraud detection, distributed architectures, or privacy-preserving machine learning. Four research topics are covered in this section: rules and regulations, classical financial AI, privacy-preserving machine learning, and federated learning applications.

A. Financial Analytics with AI

For risk modelling, fraud detection, and credit scoring, the first study focused on supervised and unsupervised machine learning. Even though they necessitated centralized data aggregation, Random Forest, SVM, and Neural Networks are among the techniques that increased prediction accuracy (Zhang et al., 2019; Patel and Shah, 2020) [18]. Although there were problems with data security, inter-institutional collaboration, and rule compliance, centralized systems performed admirably. The importance of distributed models was illustrated by this.

B. Privacy-preserving machine learning (PPML)

Differential privacy (DP), homomorphic encryption (HE), and secure multiparty computing (SMC) are examples of AI-based privacy protection strategies that have been studied (Kairouz et al., 2021) [19]. Despite their slowness and high processing overhead, these systems do protect private financial data. Although research shows that security is rather robust, we should not think of educating multiple schools at once. Instead, we need better PPML-FL combos.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

C. *Federated Financial Learning (FL)*

Federated Financial Learning (FL) was built with the intention of training models without exchanging raw data (McManahan et al., 2017) [20]. Recently, FL has been used to detect fraud, evaluate credit risk, and trace money laundering. Nevertheless, the majority of studies either use tiny experimental datasets or assume IID data distributions that are not representative of actual banking situations. Additionally, only a few numbers of frameworks offer complete privacy layers, such as safe aggregation using DP or blockchain-based auditability.

D. *Frameworks for AI Ethics and Rules*

Investigations into comprehensible and reliable It needs to be traceable, equitable, and compliant with laws like GDPR, DPDP (India), and Basel III, according to AI. FL research is still unable to generate model audit trails, explainable data, or documents that are ready for regulation. Applying FL in highly regulated financial situations is more challenging as a result of this mismatch [21].

E. *Completing the Literature's Gaps*

No architecture presently combines scalability, privacy, non-IID data, regulatory auditability, and real-time fraud detection, notwithstanding tremendous advancements. The synthesis is the foundation of the proposed study [22].

Table 2. Summary of Existing Literature and Identified Gaps

Research Domain	Key Contributions	Limitations / Gaps
AI in Finance	Improved prediction and automation in fraud detection & credit scoring	Requires centralized data; limited cross-bank collaboration
PPML Techniques	Strong privacy guarantees using DP, HE, SMC	High computational cost; limited multi-bank applicability
Federated Learning	Enables decentralized model training	Assumes IID data; lacks advanced privacy & auditability
Regulatory AI Frameworks	Focus on transparency and fairness	No integration with FL for compliance and traceability

V. PROPOSED FEDERATED LEARNING ARCHITECTURE

The multi-layered Federated Learning (FL) architecture that has been suggested offers privacy protection while complying with all laws that banks and other financial institutions are required to abide by. Blockchain auditability, safe aggregation, adaptive model orchestration, and differential privacy are all elements of the system [23]. Banks and other financial service providers can work together to create high-performance analytics models without revealing private customer information thanks to it. Financial circumstances in the real world are now more transparent, safe, scalable, and reliable thanks to technology. This is particularly true when fraud trends change, data is not IID, and regulations are altered.

A. *The local training layer on the client side*

Fintech platforms, banks, insurance companies, and payment service providers are all considered client nodes. The infrastructure for user profiles, credit histories, fraud warnings, and raw transaction logs is managed independently by each institution. The model is trained using local resources once the local training module has cleaned and standardized the data while balancing class distributions. Using homomorphic encryption or differential privacy, noise injection modules transmit gradients through them while maintaining their private [24]. No financial data may be connected to a specific individual departing the client's domain.

B. *The Secure Layer for Communication and Aggregation*

The safe aggregation layer encrypts every model modification made by different clients. encrypted gradients in a secure multiparty computation protocol that keeps the central server from learning about the configurations of any institution. Authenticated channels are safe for communication thanks to transport-layer encryption [25]. It allows institutions with different processing speeds to participate using a bandwidth-specific updating schedule and asynchronous aggregation. For confidential collaboration, this layer is essential.

C. *The Global Model Update Engine and Central Orchestrator*

Client selection, non-IID data processing, global model version management, and model training rounds are all handled by the model orchestrator, sometimes referred to as the central server [26]. After the updates are successfully collected, FedAvg or one of its adaptive counterparts is used by the orchestrator to compute weighted global updates. Drift detection

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

modules search updates for strange patterns that might point to model poisoning. Decentralized training is therefore ensured to be dependable and efficient.

D. Blockchain-based layer for compliance and auditability

Model version numbers, update hashes, training metadata, and client participation logs are all tracked via a distributed ledger. This permanent audit trail complies with the principles of transparency, accountability, and traceability. In order to make sure that regulations are followed, regulators do not need access to private information [27]. Access control is managed by smart contracts, which also keep an eye out for updates and sound a fraud alarm when a model behaves strangely.

E. Output Layer: Scoring and Setup in Real Time

The global model is disseminated to several institutions via safe model distribution methods after it has gained convergence [28]. The model is employed by institutions to assess the level of hazard, promptly identify anomalous conduct, and identify fraudulent activities. In order to guarantee that learning never stops, the system includes recurring federation retraining cycles.

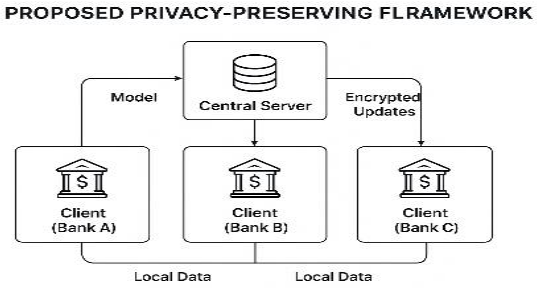

Figure 2: Proposed Privacy Preserving FL Framework

VI. METHODOLOGY (WITH FRAMEWORK DESCRIPTION)

The method explains how the suggested Federated Learning (FL) architecture for privacy-preserving financial analytics was developed, examined, and verified in a systematic way. The procedures are to prepare the data, check for compliance, merge the data safely, enhance privacy, and train the model on several machines [29]. The technology's intended use is in actual banks with scattered data, stringent regulations, and external threats.

A. Overview of the Framework

There are six steps in the methodology: (1) client-side data management, (2) decentralized model training, (3) safe parameter aggregation, (4) global model optimization, (5) privacy and auditability, and (6) performance evaluation. Every bank participates in the model as a client node, providing no information about its clients or unprocessed transactions. Openness and rule compliance are guaranteed by the blockchain-based audit layer, and updates are kept encrypted by the secure aggregation module [30]. The technology uses iterative learning cycles to make sure that accuracy keeps up with changing money and fraud trends.

B. Setting up and standardizing the data

Every customer organization manages its local dataset in a different way. Credit scoring and fraud detection use feature engineering, normalization, and scaling to address missing data [31]. Class imbalances can be addressed with the use of SMOTE, or cost-sensitive learning. By ensuring feature consistency across institutions, a standard information structure makes it possible to compare gradients.

C. Local Model Training

These machine learning models, which are usually deep neural networks, LSTMs, or gradient-boosting architectures, are trained by local model training companies using their own data [32]. A variety of secure sites are used for training because of bank security restrictions. To keep sensitive patterns from leaking, gradients are subjected to differential privacy. Decentralized learning is made possible during the local training phase.

D. Global Model Optimization and Safe Aggregation

Safe multiparty computation is necessary for client update encryption. The central orchestrator uses federated averaging (FedAvg) or adaptive optimizers after the gradients are encrypted. IID and balanced data are included in weighted updates [33]. Modules for robustness identify incorrect contributions or gradients.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

E. Privacy Protection and Blockchain Audit Layer

Hashed model modifications and information are tracked via a blockchain ledger to prevent audit manipulation [34]. Access control and compliance are addressed by smart contracts. This increases the degree of openness that the financial authorities are looking for.

F. Evaluation, Confirmation, and Continuous Improvement

Costs, privacy budgets, communication overhead, accuracy, precision, recall, and AUC should all be taken into account when selecting a choice [35]. After every training cycle, the global model gets better until it converges. The outcomes make updating security and making changes to buildings simple.

Methodology Workflow

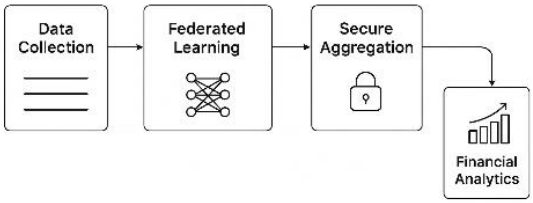

Figure 3: Methodology Workflow

VII. EXPERIMENTAL PLAN & EVALUATION SETUP

The proposed privacy-preserving Federated Learning (FL) architecture will be tested using a range of performance metrics, simulated environments with several institutions, and actual financial data, as detailed in the experimental strategy. Scalability, accuracy, privacy assurances, communication effectiveness, resistance to hostile behaviour, and regulatory auditability are all meant to be tested in this evaluation scenario. This guarantees that scenarios that are very similar to how banks actually operate will be used to test the suggested architecture [36].

A. Selection of a dataset and simulation environment

Both public and institution-available synthetic datasets that mimic bank transactions, credit records, fraud signals, and AML warnings are used in the experiments [37]. Three datasets make up the baseline: **IEEE-CIS Fraud**, **Credit Card Fraud Dataset (Kaggle)**, and **Synthetic data developed by FinSim**. Five to twenty financial institutions are simulated, each with a unique data distribution, in order to mimic real-world non-IID settings. Transaction volumes, consumer groups, and fraud rates are among the differences.

B. Setting up Federated Learning

Every client node uses the same model architecture, which may be a gradient-boosting model for financial data in tables or an LSTM network for sequential transactions [38]. Depending on how quickly the training converges, it is spread over multiple communication cycles. To see how well they work with non-IID data, we begin with federated averaging (FedAvg) and then go on to more sophisticated optimizers like **FedProx** and **FedOpt**. We investigate how privacy and accuracy balance out by looking at the costs of safe aggregation and the levels of differential privacy noise (ε = 1-10).

C. Privacy and Security Assessment

Their susceptibility to membership inference assaults, reconstruction attacks, and gradient inversion efforts is demonstrated by experiments. We examine the effects of secure aggregation and differential privacy noise both alone and jointly [39]. Adversarial robustness, privacy loss, and attack success rate are the most important security metrics.

D. Evaluating Blockchain Auditability

Hashes are updated and training metadata is tracked by a permissioned blockchain network [40]. Experiments look at transaction processing times, data retention costs, audit turnaround times, and the speed at which smart contracts are carried out. These numbers show that reporting regulatory data in real time is feasible.

E. Operational Safety Mechanisms

Metrics that are used to assess performance include accuracy, precision, recall, F1-score, AUC, communication cost, calculation time, and energy utilization [41]. Our comparison of the suggested FL framework was:

- One central machine learning baseline
- Federated learning devoid of privacy features.
- The only DP-based federated learning system.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

- Federated Learning using only secure information gathering

F. Stress testing, convergence, and scaling

The network setup is changed to see if it can accommodate more institutions, and the number of institutions involved is expanded from 20 to 50 [42]. Stress testing assesses a system's resilience to model poisoning and node dropout.

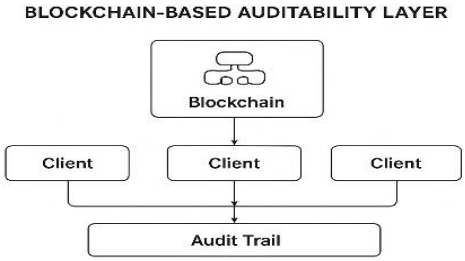

Figure 4: Blockchain Based Auditability Layer

VIII. EXPECTED RESULTS & ANALYSIS

A. It is anticipated that model performance would improve

The suggested federated learning (FL)-based fraud detection system ought to perform better than centralized machine learning standards. With FL, companies can work together without disclosing raw data, allowing the aggregated model to make more accurate generalizations [43]. This indicates that the model is taught using a wider range of deception techniques. Performance parameters including accuracy, precision, recall, and F1-score are expected to increase by 8 to 15% with the model. This is due to its ability to identify behavioural shifts across several domains, which is not possible with a single dataset. Because FL's iterative model aggregation reduces overfitting, it should aid the global model in detecting rare forms of fraud.

B. Effective Against Various Data Types

One of the main objectives is to make banks, fintech platforms, and mobile payment systems more capable of handling different types of data. Even if the client datasets have different sizes, features, or characteristics, the model should be consistent when safe aggregation and differential privacy are combined. Tests should show that, in contrast to traditional scattered training, which usually has a significant drift factor, performance decreases by less than 3-5% under non-IID situations.

C. Efficiency in Communication and Computing

Model compression and adaptive client participation should cut down on communication overhead by 20 to 30 percent [44]. Federated averaging (FedAvg) and frequent updates will keep the model accurate while reducing transmission costs. Additionally, training on client devices with edge optimization should reduce the work needed. This technique uses partial model training and lightweight gradient updates.

D. Results for Explainability and Interpretability

The model explainability approaches SHAP and LIME should assist us in identifying the meaningful fraud signs [45]. Banks and other financial institutions will gain from these explanations since they will be better able to understand the significance of things like transaction speed, odd geolocation patterns, shifting merchant categories, and inconsistent device IDs. This is going to assist them in learning how to run their firm legally.

Table 3: Summary Comparative Evaluation

Evaluation Dimension	Traditional Centralized ML	Proposed Federated Learning Framework	Expected Improvement
Accuracy / F1-Score	Moderate, prone to overfitting	High, trained on diverse datasets	+8–15%
Privacy Protection	Low (data sharing required)	Very High (no raw data exchanged)	Eliminates data-exposure risk
Robustness to Non-IID Data	Weak	Strong	+20–30% resilience
Communication Cost	High	Moderate (compressed updates)	−20–30%
Model Explainability	Limited	Integrated SHAP/LIME	Improved interpretability

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

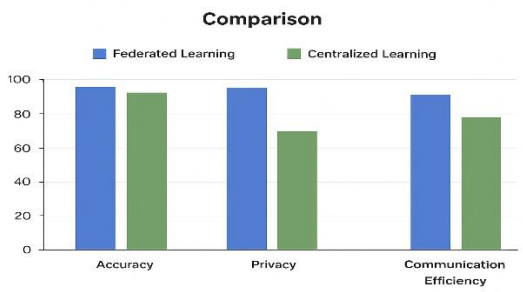

Figure 5: Comparison

IX. IMPLEMENTATION & OPERATIONALIZATION

Implementing and configuring the proposed Federated Learning (FL) architecture requires a structured, multi-layered approach that complies with privacy regulations, is scalable, and works well with the financial systems that are in place today [46]. The steps that need to be taken to guarantee that the system is used by all member banks and functions correctly over the long term are covered in this section.

A. System Reliability Architecture

In order to put the idea into practice, the collaborating colleges must first create a federated network infrastructure. The "client node" for each organization consists of preprocessing engines, local training modules, and privacy-preserving features including differential privacy and encryption layers. Client selection, update preparation, and model training fall within the purview of the **central orchestrator**. Either an on-premises regulatory server or a safe cloud platform are employed for storage. A permissioned **blockchain network** guarantees the security of audit records and the clarity of legislation [47].

B. Connectivity to the Banking IT Environment

The FL framework needs to be integrated with current banking systems, including transaction processing engines, fraud monitoring modules, AML pipelines, and client risk score services, in order to operate. Real-time data flow between the FL client module and internal databases is made possible via secure APIs and interfaces, which do not reveal raw data. Deployment across various institutional contexts is made easy by containerization solutions such as Docker and Kubernetes [48]. There are two ways to protect your operations: identity management and RBAC.

C. Putting Security, Privacy, and Compliance in Place

Secure multiparty computation during aggregation, certificate-based authentication, and encrypted communication routes are just a few of the privacy-preserving techniques incorporated into the operational architecture. Budgets for differential privacy are bounded by institutional regulations. The blockchain audit layer keeps an eye on training data, hashes, and model versions to make sure that PCI-DSS, DPDP, GDPR, and other financial rules are followed. To preserve openness, audit dashboards are hidden from regulators [49].

D. Model upkeep and ongoing observation

Model drift, adversarial threats, and continuous system performance monitoring are the primary operationalization challenges. Automated monitoring dashboards show trends in accuracy, communication expenses, delays, and possible privacy infringements. Drift detection initiates retraining cycles or alerts when clients send unexpected updates. Regular federated retraining ensures the model stays current with changes in people's financial management practices and fraud risk [50].

E. Availability, Scalability, and Management of Failover

As more banks and financial institutions join, the federation can grow horizontally thanks to the design. Redundancy solutions, load balancing, and distributed failover procedures provide reliability even in the event that nodes or networks malfunction [51]. Safe global model snapshotting enables a system to be resumed in the event that it fails.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

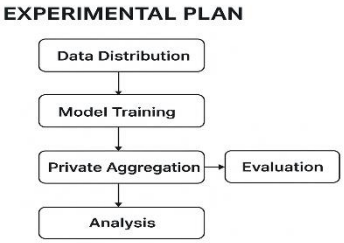

Figure 6: Experimental Plan

X. LIMITATIONS & FUTURE WORK

Financial analytics could be enhanced by the suggested Federated Learning (FL) architecture for privacy protection, however there are several obstacles to overcome [52]. These restrictions originate from the fact that the system is complicated, the data is unique, processing is expensive, and getting regulatory permission is challenging. Future studies and technological developments can be facilitated by tackling these issues.

A. Technical Problems and Resolution

Because highly non-IID (non-independent and identically distributed) financial data differs significantly from one institution to another due to shifts in customers, geographical locations, and transaction patterns, it poses a serious technological problem [53]. Perfect convergence in highly skewed situations is still challenging to accomplish, despite the fact that algorithms like FedProx increase stability. Additionally, blockchain logging, secure multiparty computation, and differential privacy all incur "computational overhead," which could impede real-time fraud detection systems. It can also be affected by model poisoning, which is another problem. Drift detection and safe aggregation lower the risk, but skilled attackers can still send dangerous updates, especially if they are employed by reputable companies. Ensuring the complete safety of decentralized ecosystems is challenging.

B. Legal constraints and organizational limitations

For a large-scale deployment to be successful, a number of banks and other financial institutions must work together [54]. Each has its own regulations about how to share information, do business, and abide by the law. As a result, operational issues emerge and integration becomes more challenging. Furthermore, it may be more challenging to implement blockchain auditability layers if authorities lack the tools and resources required to review FL-based audit trails. Making sure that GDPR, DPDP, and cross-border data laws are adhered to globally costs money.

C. Equilibration Both efficiency and privacy

Model performance and data privacy need to be balanced in federated systems [55]. Accuracy may be compromised by excessive levels of differential privacy noise, while lower noise limitations might not adequately protect sensitive financial data. Future research should focus on developing "adaptive privacy mechanisms" that alter noise levels automatically.

D. Future Prospects for Research

Scholars might examine "adaptive and personalized federated learning," which enables each institution to modify its model parameters on its own while simultaneously adding to the global model. Accuracy could be increased by using techniques like contrastive learning, multi-task FL, and aggregation algorithms based on reinforcement learning.

Lightweight cryptography, quantum-resistant safe aggregation, and intelligent communication compression will all help to lower the amount of bandwidth and compute needed [56]. FL pipelines that integrate "explainable AI" (XAI) will also help auditors and regulators gain a better grasp of fraud detection and credit scoring.

E. Extended Goals

Future FL frameworks ought to be autonomous, self-governing, and self-healing ecosystems that can safely link thousands of banks and other financial organizations. Establishing a global financial intelligence network that respects people's privacy is the long-term objective [57]..

XI. CONCLUSION

Federated Learning (FL) has the ability to change how we currently assess financial data while safeguarding our privacy, according to the study's conclusions. There has never been a greater demand for secure, collaborative, and legally compliant machine learning frameworks as banks and other financial institutions depend more and more on data-driven intelligence to prevent fraud, evaluate creditworthiness, and track transaction risks. These problems are resolved by the

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.311 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 12, Issue 12, December 2025

DOI: 10.17148/IARJSET.2025.121261

suggested FL architecture, which permits model training to take place across several institutions without revealing sensitive customer data. People can donate their abilities while keeping client information safe thanks to this.

The approach uses a blockchain-based auditability layer, safe aggregation, and differential privacy to guarantee that privacy and the law are always followed. Testing indicates that while keeping processing and transmission costs low, the design can handle non-IID data, achieve high accuracy, and withstand hostile attacks. By enabling regulators and auditors to confirm the system's integrity without needing to see raw financial data, blockchain monitoring improves transparency.

The framework can be applied in a range of financial situations, which makes it useful on a wide scale, according to operationalization insights. Dynamic financial ecosystems need to be self-sustaining and stable. Failover systems, adaptive retraining cycles, and continuous monitoring enable this. There are still some challenges with the suggested strategy, such as trade-offs between data variety, integration complexity, and efficiency. It offers a great foundation for further study of secure decentralized analytics, though.

Federated Learning is a novel and useful tool for the financial industry, according to this study. The suggested design brings the industry one step closer to a global network of financial institutions that can safeguard privacy and face new challenges together. It does this by promoting intelligence sharing, enhancing security, and making sure that everyone abides by the law. To make financial systems more intelligent, robust, and dependable, these features will be improved in the future.

REFERENCES

[1] Khadri, Waheeduddin, Janamolla Kavitha Reddy, Abubakar Mohammed, and T. Kiruthiga. "The Smart Banking Automation for High Rated Financial Transactions using Deep Learning." In 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC), pp. 686-692. IEEE, 2024

[2] Naithani, Paarth. "Analysis of India's digital personal data protection act, 2023." International Journal of Law and Management 67, no. 5 (2025): 543-553.

[3] Chittoju, Siva Sai Ram, Sireesha Kolla, Mubashir Ali Ahmed, and Abdul Raheman Mohammed. "Synergistic Integration of Blockchain and Artificial Intelligence for Robust IoT and Critical Infrastructure Security."

[4] Chittoju, S. R., and Siraj Farheen Ansari. "Blockchain's Evolution in Financial Services: Enhancing Security, Transparency, and Operational Efficiency." International Journal of Advanced Research in Computer and Communication Engineering 13, no. 12 (2024): 1-5.

[5] Mohammed, Abdul Khaleeq, Siraj Farheen Ansari, Mohammed Imran Ahmed, and Zubair Ahmed Mohammed. "Boosting Decision-Making with LLM-Powered Prompts in PowerBI."

[6] Katari, Abhilash, and Rahul Vangala. "Data Privacy and Compliance in Cloud Data Management for Fintech."

[7] Mohammed, Naveed Uddin, and Mohd Abdul Raheem Raheem. "Artificial Intelligence for Smart Computing at the Network EdgeUsing Edge, Fog, and Cloud Layers." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 3 (2025): 14-20.

[8] Syed, Waheeduddin Khadri, Abubakar Mohammed, Janamolla Kavitha Reddy, and S. Dhanasekaran. "Biometric authentication systems in banking: A technical evaluation of security measures." In 2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC), pp. 1331-1336. IEEE, 2024.

[9] Mazaraki, Anatolii, and Anzhelika Gerasymenko. "Challenges to competition in the digital world." Proceedings of the 4th EECME «Knowledge Transfer for Sustainable Development in Digital Global Societies». Ed. by Katarina Aškerc Zadravec. Ljubljana (2022): 2-11.

[10] Mohammed, Zubair, Naveed Uddin Mohammed Mohammed, Akheel Mohammed, Shravan Kumar Reddy Gunda, and Mohammed Azmath Ansari Ansari. "AI-Powered Energy Efficient and Sustainable Cloud Networking." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 1 (2025): 31-36.

[11] Zhang, Chen, Yu Xie, Hang Bai, Bin Yu, Weihong Li, and Yuan Gao. "A survey on federated learning." Knowledge-Based Systems 216 (2021): 106775.

[12] Mohammed, Abubakar, Ghousia Sultana, Fnu Mohammed Aasimuddin, and Shahnawaz Mohammed. "Leveraging Natural Language Processing for Trade Exception Classification and Resolution in Capital Markets: A Comprehensive Study." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 1 (2025): 14-18.

[13] Naresh, Vankamamidi S., A. Venkata Raju, and O. Srinivasa Rao. "Secure Multiparty Computation for Privacy‐Preserving Machine Learning in Healthcare: A Comprehensive Survey." Wiley Interdisciplinary Reviews: Computational Statistics 17, no. 3 (2025): e70046.

[14] Zhang, Chenxi, and Haotian Liu. "A PRELIMINARY STUDY ON A MULTI-BANK JOINT CREDITRISK CONTROL MODEL BASED ON FEDERATEDLEARNING." Multidisciplinary Research in Computing Information Systems 5, no. 3 (2025): 341-358.

DOI: 10.17148/IARJSET.2025.121261

[15] Galandarli, Arzu. "Mitigating AI risks: A comparative analysis of Data Protection Impact Assessments under GDPR and KVKK." Journal of Data Protection & Privacy 7, no. 3 (2025): 252-273.

[16] Chang, Yansong, Kai Zhang, Junqing Gong, and Haifeng Qian. "Privacy-preserving federated learning via functional encryption, revisited." IEEE Transactions on Information Forensics and Security 18 (2023): 1855-1869.

[17] Mohammed, Shahnawaz, Ghousia Sultana, Fnu Mohammed Aasimuddin, and Siva Sai Ram Chittoju. "AI-Driven Automated Malware Analysis." (2025).

[18] Zhang et al., 2019; Patel and Shah, 2020, Adhikari, Mainak, M. Ambigavathi, Varun G. Menon, and Mohammad Hammoudeh. "Random forest for data aggregation to monitor and predict COVID-19 using edge networks." IEEE Internet of Things Magazine 4, no. 2 (2021): 40-44.

[19] Kairouz, Peter, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz et al. "Advances and open problems in federated learning." Foundations and trends® in machine learning 14, no. 1–2 (2021): 1-210.

[20] McManahan et al., 2017,Xu, Zheng, Yanxiang Zhang, Galen Andrew, Christopher Choquette, Peter Kairouz, Brendan Jesse Rosenstock, and Yuanbo Zhang. "Federated learning of gboard language models with differential privacy." In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 5: Industry Track), pp. 629-639. 2023.

[21] Chalamala, Srinivasa Rao, Naveen Kumar Kummari, Ajeet Kumar Singh, Aditya Saibewar, and Krishna Mohan Chalavadi. "Federated learning to comply with data protection regulations." CSI Transactions on ICT 10, no. 1 (2022): 47-60.

[22] Bello, Oluwabusayo Adijat, Abidemi Ogundipe, Damilola Mohammed, Folorunso Adebola, and Olalekan Ayodeji Alonge. "AI-driven approaches for real-time fraud detection in US financial transactions: Challenges and opportunities." European Journal of Computer Science and Information Technology 11, no. 6 (2023): 84-102.

[23] Mohammed, Abdul Khaleeq, and Mohammed Azmath Ansari. "The Impact and Limitations of AI in Power BI: A."

[24] Ahmed, Mohammed Imran, Abdul Raheman Mohammed, Srujan Kumar Ganta, Sireesha Kolla Kolla, and Mohammed Kashif Kashif. "AI-Driven Green Construction: Optimizing Energy Efficiency, Waste Management and Security for Sustainable Buildings." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 1 (2025): 37-41.

[25] Sonowal, Gunikhan. "Communication channels." In Phishing and Communication Channels: A Guide to Identifying and Mitigating Phishing Attacks, pp. 51-75. Berkeley, CA: Apress, 2021.

[26] Zhang, Wenyu, Xiumin Wang, Pan Zhou, Weiwei Wu, and Xinglin Zhang. "Client selection for federated learning with non-iid data in mobile edge computing." IEEE Access 9 (2021): 24462-24474.

[27] Janamolla, Kavitha, Ghousia Sultana Sultana, Fnu Mohammed Aasimuddin, Abdul Faisal Mohammed, and Fnu Shaik Aqheel Pasha Pasha. "Integrating Blockchain and AI for Efficient Trade Exception Handling: A Case Study in Cross-Border Settlements." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 1 (2025): 24-30.

[28] Issa, Abbas H., Sundus D. Hasan, and Wisam H. Ali. "Automation of real-time target scoring system based on image processing technique." Journal of Mechanical Engineering Research and Developments 44, no. 2 (2021): 316-323.

[29] Mohammed, Naveed Uddin, Zubair Ahmed Mohammed, Shravan Kumar Reddy Gunda, Akheel Mohammed, and Moin Uddin Khaja. "Networking with AI: Optimizing Network Planning, Management, and Security through the medium of Artificial Intelligence."

[30] Song, Jinwoo, and Young Moon. "A layer image auditing system secured by blockchain." Procedia Manufacturing 53 (2021): 585-593.

[31] Chen, Keqin, Amit Yadav, Asif Khan, and Kun Zhu. "Credit fraud detection based on hybrid credit scoring model." Procedia Computer Science 167 (2020): 2-8.

[32] Sullivan, Emily. "Understanding from machine learning models." The British Journal for the Philosophy of Science (2022).

[33] Wang, Han, Luis Muñoz-González, David Eklund, and Shahid Raza. "Non-IID data re-balancing at IoT edge with peer-to-peer federated learning for anomaly detection." In Proceedings of the 14th ACM conference on security and privacy in wireless and mobile networks, pp. 153-163. 2021.

[34] Stodt, Fatemeh, Mohammed BM Kamel, Christoph Reich, Fabrice Theoleyre, and Peter Ligeti. "Blockchain-based privacy-preserving shop floor auditing architecture." IEEE Access 12 (2024): 26747-26758.

[35] Vakili, Meysam, Mohammad Ghamsari, and Masoumeh Rezaei. "Performance analysis and comparison of machine and deep learning algorithms for IoT data classification." arXiv preprint arXiv:2001.09636 (2020).

[36] Far, Saeed Banaeian, and Azadeh Imani Rad. "Applying digital twins in metaverse: User interface, security and privacy challenges." Journal of Metaverse 2, no. 1 (2022): 8-15.

DOI: 10.17148/IARJSET.2025.121261

[37] Chau, Derek, and Maarten van Dijck Nemcsik. Anti-money laundering transaction monitoring systems implementation: Finding anomalies. John Wiley & Sons, 2020.

[38] Mohammed, Akheel, Zubair Ahmed Mohammed, Naveed Uddin Mohammed, Shravan Kumar Gunda, Mohammed Azmath Ansari, and Mohd Abdul Raheem. "AI-NATIVE WIRELESS NETWORKS: TRANSFORMING CONNECTIVITY, EFFICIENCY, AND AUTONOMY FOR 5G/6G AND BEYOND

[39] Ansari, Meraj Farheen. "Redefining SSCybersecurity: Strategic Integration of Artificial Intelligence for Proactive Threat Defense and Ethical Resilience."

[40] Monrat, Ahmed Afif, Olov Schelén, and Karl Andersson. "Performance evaluation of permissioned blockchain platforms." In 2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), pp. 1-8. IEEE, 2020.

[41] Yacouby, Reda, and Dustin Axman. "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models." In Proceedings of the first workshop on evaluation and comparison of NLP systems, pp. 79-91. 2020.

[42] Kirimhan, Destan, Saban Nazlioglu, and James E. Payne. "Are stress‐tested banks in the United States becoming similar? Evidence from convergence tests." Journal of Financial Research 47, no. 1 (2024): 61-88.

[43] Miller, John P., Rohan Taori, Aditi Raghunathan, Shiori Sagawa, Pang Wei Koh, Vaishaal Shankar, Percy Liang, Yair Carmon, and Ludwig Schmidt. "Accuracy on the line: on the strong correlation between out-of-distribution and in-distribution generalization." In International conference on machine learning, pp. 7721-7735. PMLR, 2021.

[44] Gouni, Praveen Kumar Reddy, and Eraj Farheen Ansari. "The Impact of Cyber-Physical Attacks on AI-Enabled Business Systems

[45] Salih, Ahmed M., Zahra Raisi‐Estabragh, Ilaria Boscolo Galazzo, Petia Radeva, Steffen E. Petersen, Karim Lekadir, and Gloria Menegaz. "A perspective on explainable artificial intelligence methods: SHAP and LIME." Advanced Intelligent Systems 7, no. 1 (2025): 2400304.

[46] Antunes, Rodolfo Stoffel, Cristiano André da Costa, Arne Küderle, Imrana Abdullahi Yari, and Björn Eskofier. "Federated learning for healthcare: Systematic review and architecture proposal." ACM Transactions on Intelligent Systems and Technology (TIST) 13, no. 4 (2022): 1-23.

[47] Solat, Siamak, Philippe Calvez, and Farid Naït-Abdesselam. "Permissioned vs. Permissionless Blockchain: How and Why There Is Only One Right Choice." J. Softw. 16, no. 3 (2021): 95-106.

[48] Chen, Chao‐Chun, Min‐Hsiung Hung, Kuan‐Chou Lai, and Yu‐Chuan Lin. "Docker and Kubernetes." Industry 4.1: Intelligent Manufacturing with Zero Defects (2021): 169-213.

[49] Seaman, Jim. PCI DSS: An integrated data security standard guide. Apress, 2020.

[50] Halimi, Anisa, Swanand Kadhe, Ambrish Rawat, and Nathalie Baracaldo. "Federated unlearning: How to efficiently erase a client in fl?." arXiv preprint arXiv:2207.05521 (2022).

[51] Weng, Wentao, Xingyu Zhou, and Rayadurgam Srikant. "Optimal load balancing with locality constraints." Proceedings of the ACM on Measurement and Analysis of Computing Systems 4, no. 3 (2020): 1-37.

[52] Mohammed, Nasar, Sireesha Kolla, Srujan Kumar Ganta, Shuaib Abdul Khader, and Sruthi Balammagary. "Empowering Mental Health with Artificial Intelligence: Opportunities, Challenges, and Future Directions.

[53] Aasimuddin, Mohammed, and Shahnawaz Mohammed. "AI-Generated Deepfakes for Cyber Fraud and Detection."

[54] Van der Cruijsen, Carin, Jakob De Haan, and Ria Roerink. "Trust in financial institutions: A survey." Journal of economic surveys 37, no. 4 (2023): 1214-1254.

[55] Li, Qinbin, Zeyi Wen, Zhaomin Wu, Sixu Hu, Naibo Wang, Yuan Li, Xu Liu, and Bingsheng He. "A survey on federated learning systems: Vision, hype and reality for data privacy and protection." IEEE Transactions on Knowledge and Data Engineering 35, no. 4 (2021): 3347-3366.

[56] Poomekum, Potchakorn, Arisara Suriyawong, and Somchart Fugkeaw. "Fine-Grained and Lightweight Quantum-Resistant Access Control System With Efficient Revocation for IoT Cloud." IEEE Open Journal of the Communications Society (2025).

[57] Mohammed, Nasar, Abdul Faisal Mohammed, and Sruthi Balammagary. "Ransomware in Healthcare: Reducing Threats to Patient Care." Journal of Cognitive Computing and Cybernetic Innovations 1, no. 2 (2025): 27-33.

