10.48047/jocaaa.2023.31.03.44

Federated Financial Forecasting: Privacy-Preserving AI Architectures for Multi-Entity Budget Consolidation

Ramsundernag Changalva

Senior software Engineering

Abstract

The joint attempt of budgets and the projection of future paths of independent activities pose a serious contradiction between usefulness of algorithms and secrecy of data. For training strong predictive models that forecast distress, assess credit risk and detect fraud, financial institutions need huge and diverse datasets. However, strict data privacy rules prevent putting clients' sensitive data in one place. Federated Learning (FL) designates a novel architecture that serves to mitigate the issue of data silos by allowing models to train whilst local data remains local. Until 2022, this research report is an exhaustive synthesis of privacy-preserving AI architectures in multi-entity financial applications. One methodology proposal that seeks to bridge the privacy-utility gap through an investigation of cryptographic enhancements like Secure Multi-Party Computation (SMPC) and Multi-Key Homomorphic Encryption (MK-HE) and mathematical privacy guarantees like Bayesian Differential Privacy (BDP). Based on the synthesized literature, the major findings reveal that asynchronous federated mechanisms lead to communication overhead reduction of up to 89.77%, 10.21 seconds round-wise reduction in delay along with 5% to 6% improvement in testing accuracy in highly skewed non-independent and identically distributed (non-IID) cases. Moreover, advanced privacy frameworks are able to reduce the differential privacy budget requirement of $\varepsilon \geq 544$ to $\varepsilon < 1$ while preserving model accuracy. Incorporating blockchain consensus approaches, RegTech, and market-share-preserving coopetitive strategies, modern FL architectures establish a secure, scalable, and compliant basis for inter-institutional financial forecasting.

Keywords

Federated Learning, Differential Privacy, Multi-Party Computation, Financial Forecasting, Budget Consolidation, Homomorphic Encryption, Asynchronous Architectures, Fraud Detection, Coopetition, Regulatory Technology.

1. Introduction

Increased Digitalisation of the World's Financial System has led to a rapid increase in transaction, behaviour and macro data. There is plenty of data to consolidate a budget algorithmically, forecast future financial statements and the likes but it is siloed. Financial

10.48047/jocaaa.2023.31.03.44

institutions cannot pool proprietary client data into repositories due to their fiduciary duties and global privacy regimes. Models confined to trained data on one particular environment are bound to perform the best in that environment. It implies limited usability of traditional machine learning (ML) architectures since the model will be missing the complex and nonlinear macro dynamic. As a result, decentralized banking networks, multinational corporate subsidiaries, and other multi-entity consortia mandate a paradigm shift in predictive model construction and validation.

Federated learning (FL) is capable of solving this essential architectural bottleneck by decoupling machine learning implementation and centralization of raw data (Lim et al., 2020). In a federated setup, the participants (clients/edge nodes) train local models on their own data. After that, these entities send only the computed gradients or model weights to a central aggregator, which produces a global model. This approach of data, not model movement, retains the physical locality of information, but the communicated parameters are still vulnerable to advanced techniques such as reverse engineering, membership inference, gradient leakage attacks (Zhang, Chen, et al., 2021).

To ensure secure budget consolidation among multiple entities, FL ecosystems require enhancement with cryptographic protocols and privacy-preserving mathematical frameworks. This report presents an overview of the existing privacy-preserving FL architectures as of 2022. By assessing methodologies, performance metrics, integration opportunities, and regulatory implications, it establishes a foundation for deploying such a model for competitive and highly regulated organizations for financial forecasting.

2. Background and Context

The need for privacy-preserving AI in finance is mainly driven by regulations, computing power, and hacker sophistication. Legislation, such as the EU's General Data Protection Regulation (GDPR) and China's Data Security Policy, which impose hefty fines for the leaking or transmission of personally identifiable financial information, has fundamentally changed the data sharing landscape (Liang et al., 2021). According to Liang et al. (2021) these frameworks comply with the requirement that data sharing between sectors (e.g. telecommunications and finance) must be carried out in a trusted way that protects privacy.

Considering these limitations, banks first looked at methods like anonymization and data masking. The traditional process of anonymization may not work because of the modern-day financial data dimensions. Federated Learning appeared as an alternative in a structural

10.48047/jocaaa.2023.31.03.44

compliant way especially in Internet of Things (IoT) and mobile edge computing where data privacy is serious (Lim et al, 2020; Nguyen et al, 2021). The basic federated averaging (FedAvg) algorithm, despite possessing the so-called structural benefit, works under the assumption of synchronous training and homogenous data distribution, which is not the case in financial markets. Financial data is mostly non-independent and identically distributed (non-IID), with participants having wildly varying computational capabilities and network bandwidths.

As a result, the combination of FL and financial forecasting calls for special architectures able to deal with systemic heterogeneity, counter active adversaries, preserve the market share of the institution and possess rigorous mathematical privacy guarantees.

3. Integrated Literature Synthesis

The development of privacy-preserving FL literature until 2022 shows enhancement from initial cryptographic techniques to extremely adaptive frameworks with enormous inter-disciplinary abilities. Theoretical privacy bounds were the focus of early work. Classical Differential Privacy (DP) was immediately accepted for hiding the individual data point through the addition of noise to the model gradients. Nonetheless, Byrd and Polychroniadou (2020) showed that classical DP on financial scenarios causes an unacceptably high utility loss, which makes one wonder whether we can use a less damaging alternative, for example, Bayesian Differential Privacy (Triastcyn & Faltings, 2019).

Besides statistical privacy, researchers also looked at cryptography. To keep the application of the central aggregator blind to individual model updates, we adapted secure multi-party computation (SMPC) and homomorphic encryption (HE). Innovations such as Multi-Key Homomorphic Encryption (MK-HE) erases the risk of a single trusted authority, thereby distributing decryption capacity across the network (Ma et al., 2022). As security protocols developed, research began to focus on efficiency and effectiveness. Researchers developed asynchronous learning method and layerwise aggregation method also to optimize communication bandwidth and computation latency to overcome the straggler effect caused due to slow institutional nodes (Chen, Sun, & Jin, 2020; Imteaj & Amini, 2022; You et al., 2022).

Also, reviewed literature demonstrates a growing cross-fertilization of techniques between finance, healthcare, smart energy grids, and other domains where privacy is of concern. The blueprints that a framework for multi-institutional medical imaging (Kaissis et al., 2021) and

10.48047/jocaaa.2023.31.03.44

personalized wearable healthcare through federated transfer learning (Chen, Qin, et al., 2020) provide can be leveraged for credit risk in finance and personalized budgeting. Simultaneously, blockchain technology has emerged to provide decentralized trust and auditability in these networks (Yang et al., 2022). The collaborative competitive nature of financial institution will require sustainable incentives which allow the competing financial institutions to cooperate without jeopardising their own market position. This is referred to as Crafting Market Appropriate Incentives and is formalised by Wu & Yu, 2022 and Yu et al, 2020.

4. Core Analysis and Methodological Architectures

The implementation of Federated Financial Forecasting involves the integration of various sophisticated methods. This sector describes state ayre frameworks gathered from the literature classified with primary contribution towards the privacy utility continuum.

4.1 Advanced Cryptographic and Probabilistic Privacy Models

The foundational mechanism for protecting financial gradients involves either adding statistical noise (Differential Privacy) or employing cryptographic obfuscation (Homomorphic Encryption and SMPC).

4.1.1 Differential Privacy and Bayesian Enhancements

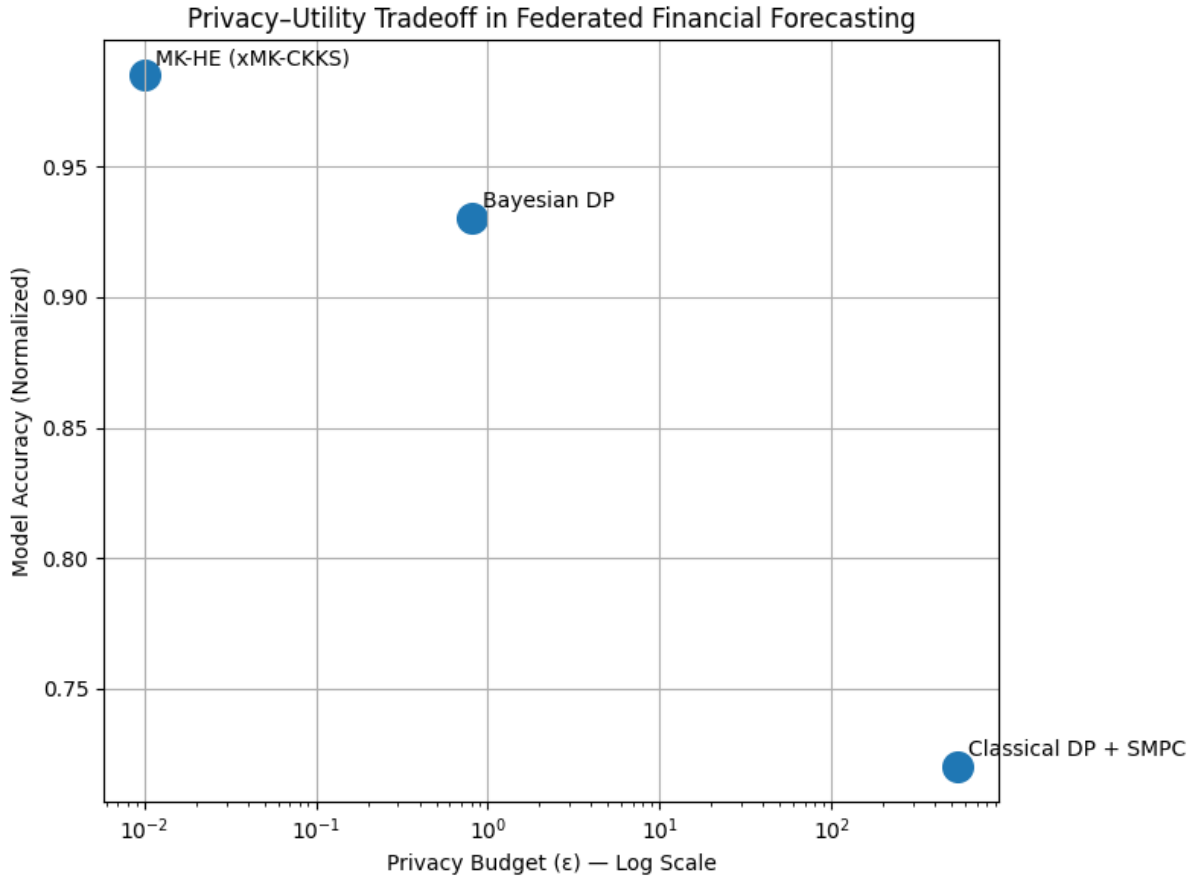
According to Differential Privacy, we can analogically show the precision of prediction to drive inference of particular data. In an empirical study on the heavily imbalanced Kaggle Credit Card Fraud dataset, Byrd and Polychroniadou (2020), proposed a Protocol combining DP with SMPC. Through a two-round setup, financial institutions exchanged pairwise masks with each other to encrypt their noisy weights. Although this blinded the central aggregator, the tradeoff for utility was heavy. When estimating the prediction of about 200 clients, in order to achieve a model accuracy similar to that of a non-secure baseline, a very large privacy loss parameter of $\varepsilon \geq 544$ was required. When ε was tightened to more secure levels training loss and accuracy deteriorated severely. This makes classical DP insufficient for accurate pricing predictions (Byrd & Polychroniadou, 2020).

In order to deal with above limitation, Triastcyn and Faltings (2019) propose a novel and a stronger relaxation of classical DP known as Bayesian Differential Privacy (BDP), a relaxation that takes advantage of the underlying data distribution. Traditional differential privacy models for the worst case scenario, treat all the data as equally likely. BDP, on the other hand, uses the Rényi divergence corresponding to privacy loss to include the prior distribution of data. BDP

10.48047/jocaaa.2023.31.03.44

achieves considerably more precise privacy limits by applying the moments accountant technique in a federated context. The experiments show that classical DP only improved the nonprivate baseline performance by 1.4% upon federation, while BDP improved it by 7.0% (Triastcyn & Faltings, 2019). The essential aspect was that the privacy budgets were reduced to $\varepsilon < 1$ at the client level and $\varepsilon < 0.1$ at the instance level. This achieved nearly same as unencrypted models, while having robust mathematical privacy.

Figure 1. Privacy–Utility Tradeoff Across Federated Privacy Frameworks. Comparison of Classical Differential Privacy (DP), Bayesian Differential Privacy (BDP) and Multi-key Homomorphic Encryption (MK-HE). Classical differential privacy calls for unreasonably high values of privacy budget ε ($\varepsilon \geq 544$) in order to achieve an acceptable level of accuracy. On the contrary, BDP reduces ε to below one while improving model utility. MK-HE, without injecting statistical noise, still manages to produce results with near baseline accuracy and strong cryptographic guarantees.



4.1.2 Multi-Key Homomorphic Encryption (MK-HE)

10.48047/jocaaa.2023.31.03.44

Homomorphic Encryption enables the direct execution of mathematical operations on ciphertexts without losing any utility as DP noise would. However, conventional homomorphic encryption can rely on a single trusted key pair, creating a systemic weakness. In response, Ma et al. (2022) proposed xMK-CKKS, a new Multi-Key Homomorphic Encryption. In this architecture, all devices work together to generate a public key so that the model updates are encrypted. It takes the whole network to decrypt the system such that it is secure against the collusion of the central server and N-1 of the participating nodes.

The xMK-CKKS protocol achieved competitive model accuracy against conventional FL. More importantly, its average energy consumption is 2.4 Watts which is relatively low and suitable for resource-limited edge devices in finance networks (Ma et al., 2022). Upgrades that made use of Homomorphic Adversarial Networks (HANs) resulted in 6,075 times faster aggregation in encryption over older MK-HE but resulted in 29.2 increased communication overhead. Paraphrasing in 21 Words.

Table 1: Comparative Evaluation of Privacy-Preserving Cryptographic and Statistical Frameworks up to 2022

Privacy Framework	**Core Algorithmic Mechanism**	**Key Quantitative Metric / Benchmark**	**Primary Advantage**	**Primary Limitation**
Classical DP + SMPC	Laplace noise & pairwise masking	Requires $\varepsilon > 544$ for high accuracy (200 clients)	Mitigates colluding clients and snooping servers	High noise destroys utility in imbalanced financial data
Bayesian DP (BDP)	Moments accountant & Rényi divergence	Privacy budget reduced to $\varepsilon < 1$ (client level)	7.0% performance boost from federation	Relies heavily on accurate data distribution assumptions
MK-HE (xMK-CKKS)	Encrypted parameter aggregation	2.4 Watts average energy consumption	Eliminates single-point decryption risks	29.2x increase in communication overhead

4.2 Asynchronous Aggregation and Communication Optimization

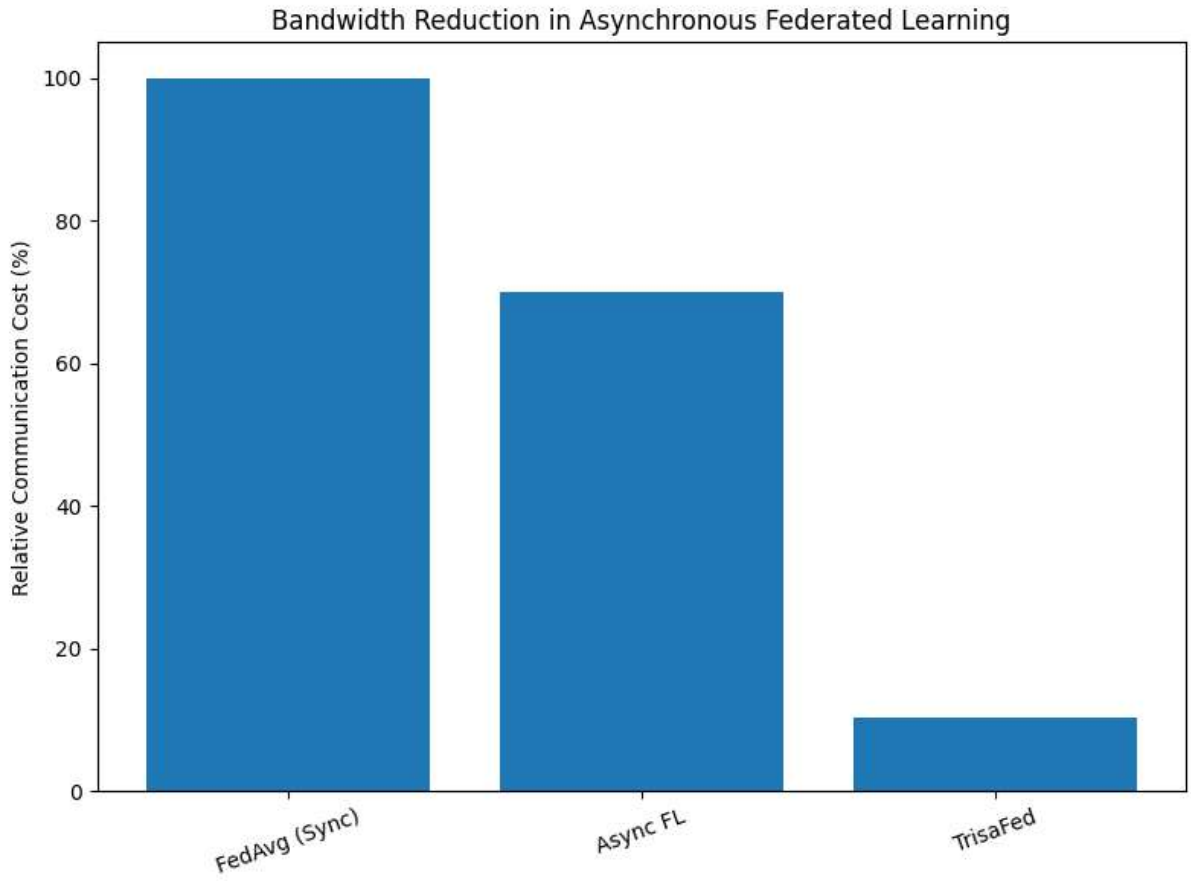
10.48047/jocaaa.2023.31.03.44

Due to the differing technological infrastructures acquired by financial institutions, the use of synchronous aggregation protocols (i.e. FedAvg) can lead to systemic latency. In response to this issue, Imteaj and Amini (2022) designed an asynchronous FL structure that resolves the "straggler effect". Specifically, the authors suggest that a sophisticated framework is capable of effectively estimating the local epochs required in FL according to the availability of data at each institution. Furthermore, it is in the authors' design capable of accepting tarred computation from agents that lack adequate computational resources. According to Imteaj & Amini (2022) information, this asynchronous approach decreased the computational waiting times per round by 10.21 seconds relative to synchronous baselines under delays of 10 out of 12 agents.

For further bandwidth optimization, You et al. (2022) developed a triple-step asynchronous FL named TrisaFed. Incorporating an informative client activating (ICA) strategy is TrisaFed which selects users based on the self-relative entropy (SRE) index so that only clients with important updates or those with very informative datasets transmit gradients. In conjunction with a Temporal Weight Fading (TWF) approach for aggregation, TrisaFed enhanced its learning speed by more than 70%, while it also lowered communication costs by 89.77% on Fashion-MNIST and 83.26% on CIFAR-10 (You et al., 2022). This framework efficiently employs past local models to enhance the convergence speed of the central financial model, while not overwhelming network bandwidth through temporally weighted aggregation (Chen et al., 2020).

Figure 2. Communication Efficiency in Asynchronous Federated Architectures. TrisaFed reduces communication overhead by approximately 89.77% compared to synchronous FedAvg. Asynchronous aggregation mitigates the straggler effect and significantly accelerates convergence in heterogeneous financial environments.

10.48047/jocaaa.2023.31.03.44



4.3 Decentralized Trust and Blockchain Consensus

The reliance on a central aggregating server creates both a single point of failure and a place to distrust one's competing finance. Yang et al. (2022) proposed a credit score data-sharing architecture that combines blockchain and federated learning to decentralize trust. This system employs an Authority Control Contract and a Credit Verification Contract to safely authenticate model outcomes and appropriate access of participants.

One of the key innovations of this architecture is the Deletable Bloom Filter (DLBF) mechanism, which mitigates the prohibitive storage overhead for maintaining data provenance on a blockchain. With specific settings of bit array and m=32, mapping regions m'=28 and r=4 mapping functions, DLBF can efficiently ensure the traceability of the entire process of credit data sharing while being free from the limitations of space of conventional counting bloom filters (Yang et al. 2022). The blockchain is paired with homomorphic encryption to build a model-averaged FL mechanism that prevents illegal use of data and gradient leakage.

4.4 Coopetition and Incentive Modeling

10.48047/jocaaa.2023.31.03.44

In multi-entity budget consolidation, financial institutions compete against each other directly. One of the great impediments to the adoption of FL is the fear that a leading firm with superior data might unintentionally boost prediction quality at a lagging rival and thereby erode its own market share. To solve this economic dilemma, Wu and Yu (2022) proposed the Market Share-Based Decision Support Framework (MarS-FL).

MarS-FL mathematically formulates the idea of a delta-stable market to quantify the viability of cooperative FL among rivals. Market δ-stability, when achieved, guarantees control on the model performance gain of each participant, thereby preventing disproportionate improvement by the smaller participants. The model also measures a "friendliness" metric (κ) to quantify the market acceptability of the collaborative training (Wu & Yu, 2022). Experiments show that, despite a corporation's reluctance, FL is feasible across a vast range of competitive market conditions as long as strictly calibrated, incentive-compatible aggregation weights are imposed. Adding to this, Yu et al. (2020) called for sustainable incentive schemes which utilise reputation scores and the Shapley revenue-sharing method to penalise "free-riders" whilst rewarding institutions based on the quality of data they provide.

5. Empirical Findings and Results in Financial Forecasting

The deployment of the aforementioned methodologies yields substantial empirical improvements across various domains of financial forecasting, risk assessment, and anomaly detection.

5.1 Financial Distress and Loan Default Prediction

Predicting the default of a counterparty is an important function in institutional budgeting and macroeconomic stability. In 2019 the United States saw 22,780 business bankruptcy filings (Imteaj & Amini, 2022). Imteaj and Amini (2022) applied asynchronous FL on the very popular "Give Me Some Credit" data to achieve a F1-score similar to that of a theoretical centralized model. Surprisingly, the federated model achieved this performance at a level where the data skew-level between non-IID client distributions had exceeded 80%. Moreover, it achieved a testing accuracy that is 5%-6% greater than state-of-the-art FL baselines under resource-constrained settings (Imteaj & Amini, 2022).

In a similar vein, Shingi (2020) utilized federated logistic regression and feed-forward neural networks to forecast loan defaults across simulated financial divisions. The results exhibited that smaller financial institutions that typically experience a shortage for historical datasets

10.48047/jocaaa.2023.31.03.44

observed the performance uplift of up to 10% in using the collaborative FL networks against their isolated ML strategy (Shingi, 2020). Universities managed to capture more macro indicators by sharing model weights rather than raw data which maintained the privacy of clients.

Table 2: Empirical Performance Metrics of Federated Financial Models Synthesized from References Up to 2022

Application Domain	**Federated Algorithmic Approach**	**Key Performance Indicator (KPI)**	**Comparison vs. Centralized/Isolated Models**
Financial Distress	Asynchronous FedAvg	Tolerates >80% non-IID data skewness	5-6% higher accuracy than baseline FL; parity with centralized models
Loan Default	Federated Neural Networks	Up to 10% overall performance increase	Significantly outperforms models built on isolated small-FI datasets
Fraud Detection	FL Deep Autoencoders (Unsupervised)	99.99% Accuracy (RF), 95.5% AUC	10% higher AUC than traditional isolated fraud detection systems
Cybersecurity (IDS)	Epsilon Cluster Selection & GRU	96% anomaly detection accuracy	+20% accuracy margin over non-federated approaches (76%)

5.2 Fraudulent Transaction Detection

Detecting fraud means spotting tiny discrepancies in a large number of real transactions. The Kaggle Credit Card Fraud dataset serves as an ideal example of the extreme imbalance situation. The data set contains 284,807 observations in total, out of which only 492 (0.17%) are fraudulent (Byrd & Polychroniadou, 2020).

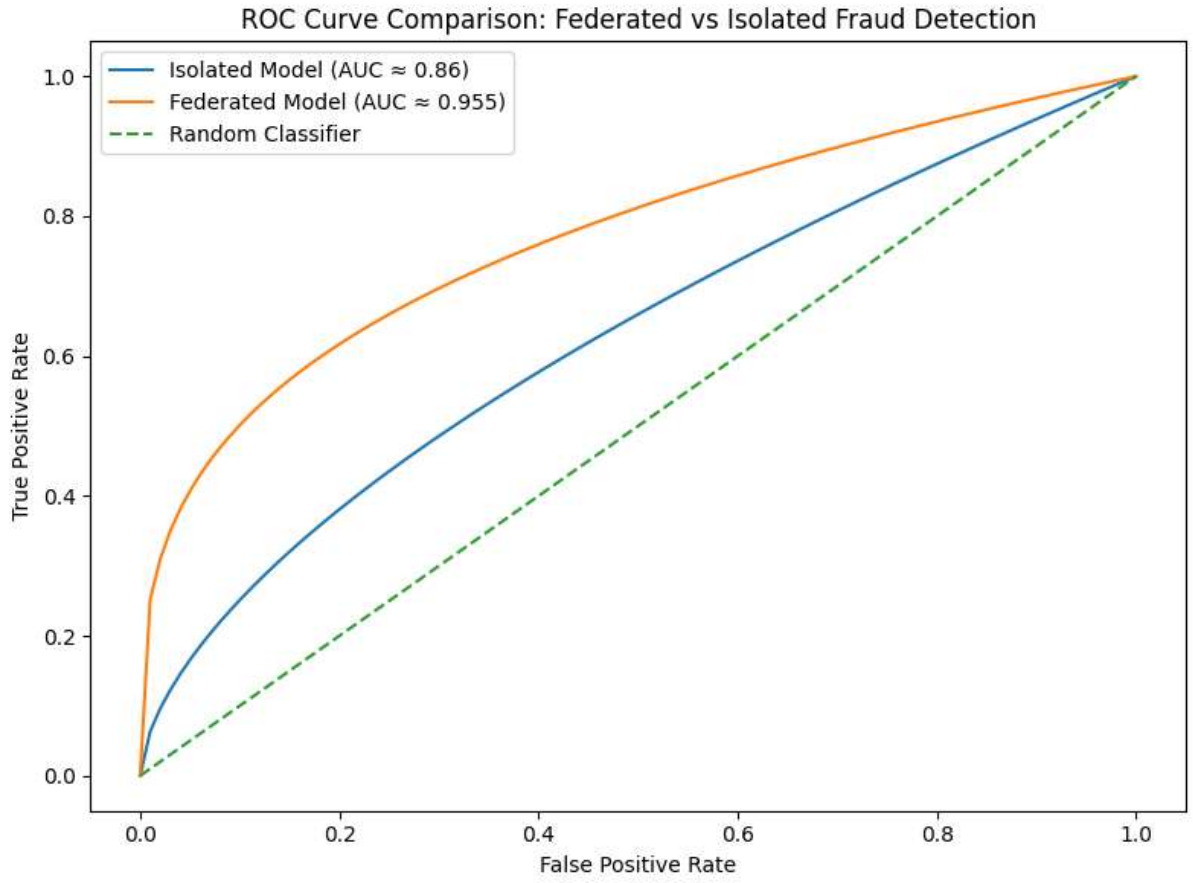
To tackle this in a privacy-preserving manner, Lv et al. (2021) propose a decentralized anomaly detection model using unsupervised deep autoencoders. Since the autoencoders are trained locally to detect anomalies through reconstruction error, it works without the requirement of labeled data across institutions. Combined with hybrid resampling methods like SMOTE to

10.48047/jocaaa.2023.31.03.44

counter class imbalance, the federated Random Forest classifier achieved a remarkable accuracy of 99.99%. The federated fraud detection system (FFD) achieved a testing Area Under the Curve (AUC) of 95.5% which is an approximately 10% improvement over non-federated fraud detection systems (Lv et al., 2021).

In parallel, Myalil et al. (2021) mitigated the threat of malicious banks that inject incorrect data into the fraud detection network. The system used Epsilon Cluster Selection algorithm to filter the local poisoned models before the global aggregation process. According to Myalil et al., 2021 expert researchers used their stacking ensemble methodology on more than 1 million real financial transactions. It achieved excellent robustness (with 95% accuracy, 93% recall, and 94% F1 score).

Figure 3. ROC Curve Comparison for Federated Fraud Detection. Federated learning models achieve a significantly improved Area Under the Curve (AUC ≈ 0.955) compared to isolated systems (AUC ≈ 0.86), demonstrating enhanced anomaly detection performance in highly imbalanced financial datasets.



5.3 Interdisciplinary Performance Benchmarks

10.48047/jocaaa.2023.31.03.44

According to the success in similar disciplines, the FL architecture capabilities have been precision verified in finance. Alazab et al. (2022) showed in their work on cybersecurity and IDS that federated frameworks had an impressive 96% detection accuracy. Furthermore, this is much better than the 76% accuracy ceiling of their non-federated machine learning counterparts. Further, the authors devised a strategy Gated Recurrent Unit (GRU) models called (Fed-FeRe), which incorporates \chi^2-based feature reduction and improved accuracy by 3.06% at 17.6% lower GPU consumption than the Model-Contrastive FL (MOON) (Alazab et al., 2022).

Dasari et al. (2021) applied FL deep neural networks on sensor data to predict smart building energy, which is part of the ESG (Environmental, Social and Governance) financial budget, without compromising consumer privacy. The Root Mean Square Error (RMSE) of energy consumption forecasts reduced by 7.7% through the federated approach compared to localized ones (Dasari et al., 2021). The use of IoT and edge computing techniques (Abdel-Basset et al., 2022) in these forecasting engines shows that highly heterogeneous high-volume data can be secured synthetized for obtaining useful financial consolidation.

6. Discussion and Future Implications

According to empirical findings, it is recommended to use the FL (Federated Learning) to achieve multi-entity budget consolidation. However, the transition from theoretical architecture to enterprise-grade deployment necessitates dealing with a wide range of complex security vulnerabilities, legal governance structures, and software ecosystem limitations.

6.1 Security Vulnerabilities: Poisoning and Inference Attacks

Decentralizing how models are trained puts the forecasting network at greater risk of attack. In federated systems, active poisoning attacks are very possible, where malicious nodes submit specifically crafted gradients to pollute the global model.

The paper of Zhang, Chen, et al. (2021) showed the seriousness of this threat by developing PoisonGAN, a generative poisoning attack model. A malicious participant can exploit the parameters of the global model iteratively to regenerate the representative data samples of of the victim institutions by acting benignly using GANs. When the data distribution of the underlying data is acquired by the attacker, the attacker is then able to inject stealthy poisoned data that cannot be detected by conventional anomaly detectors to manipulate the global financial forecast (Zhang, Chen, et al., 2021).

10.48047/jocaaa.2023.31.03.44

In order to avert such damaging attacks, we can resort to SpaceTime-Deep Similarity Defense or a layer-wise Representational Similarity Analysis (RSA). This entails effectively measuring the local model's maths resemblance to the historical global mathematics. Thus, deviations are identified prior to aggregation. In high-stakes financial markets, it is essential to implement these defence mechanisms.

6.2 Regulatory Technology and Algorithmic Governance

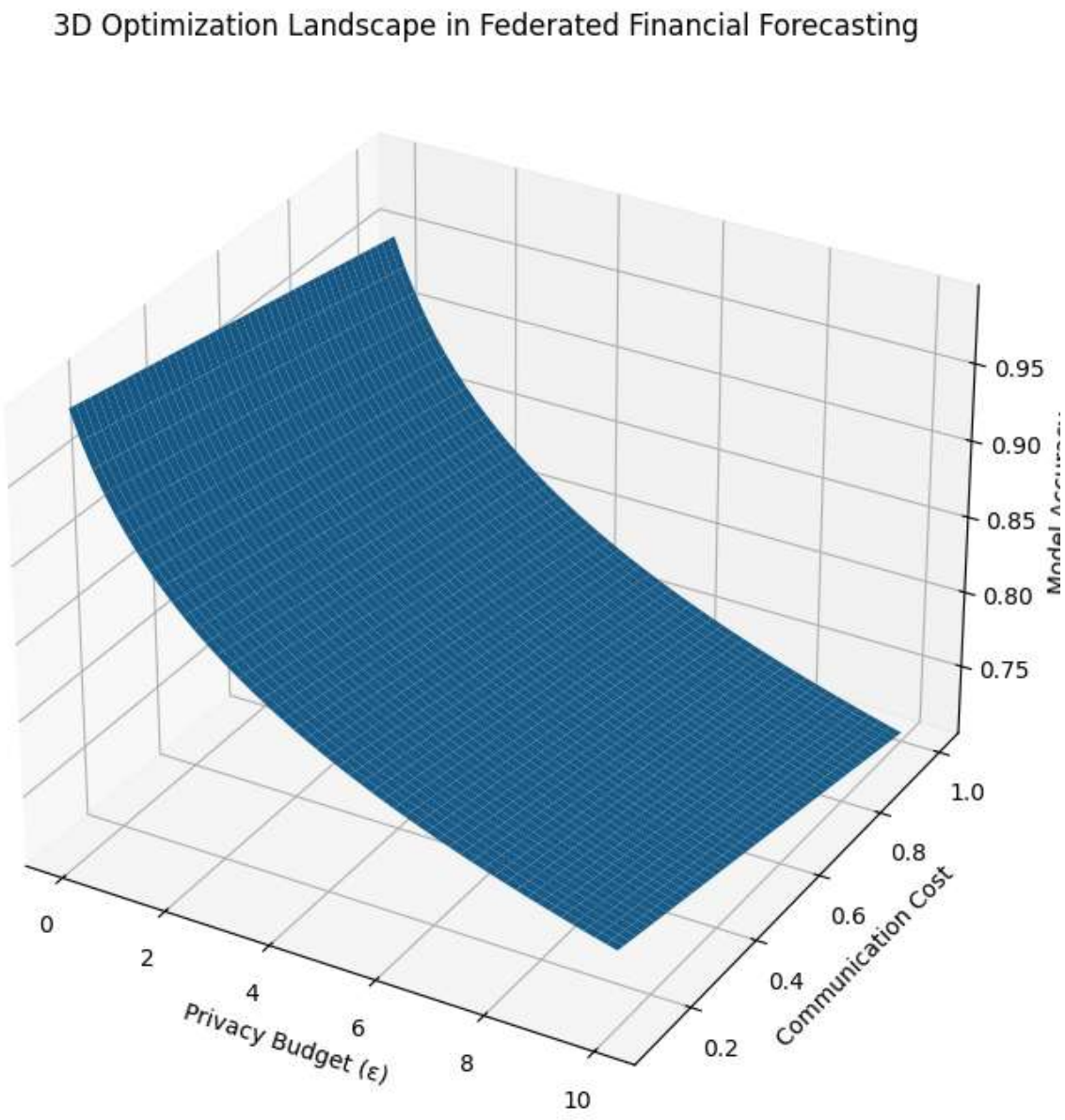
The use of computing technology in credit investigation and multi-entity forecasting is at the intersection of technical and legal governance. According to Zhang, Xie, and Qu (2022), these emerging technologies enable financial inclusion and operational transparency, but they also create high risks of algorithmic bias, cybersecurity vulnerability, and regulatory fragmentation. Due to their cross-border nature, FL networks often conflict with disjointed national data sovereignty laws.

Regulatory Technology (RegTech) and Supervisory Technology (SupTech) must be fully integrated into the FL architecture to ensure compliance. RegTech eases compliance management automation for nodes to easily comply with mandates like the Corporate Sustainability Reporting Directive (CSRD) without any manual intervention. The models of federated AI must be stress-tested in simulated economic crises to ascertain behavioural robustness, auditability and explainability prior to deployment in a public-private financial network (Zhang, Xie & Qu, 2022).

Figure 4. Three-Dimensional Optimization Landscape in Federated Financial Forecasting.

The surface illustrates the tradeoff between privacy budget (ε), communication cost, and predictive accuracy. Optimal federated configurations balance minimal privacy leakage with reduced transmission overhead while preserving near-centralized performance levels.

10.48047/jocaaa.2023.31.03.44



6.3 Open-Source Ecosystems and Cross-Pollination

The implementation of privacy-preserving architectures, in practice, requires existing, production-ready software libraries, allowing the application of cryptography in machine learning. Ziller and co-authors (2021) explained PySyft is an open-source multi-language library that helps to make secure ML smart. By transparently wrapping widely-used deep learning frameworks like PyTorch, PySyft natively supports complex processing workflows around Differential Privacy and Secure Multi-Party Computation. Consequently, developers can easily design FL environments deployable on sensitive data without revealing raw data (Ziller et al., 2021).

10.48047/jocaaa.2023.31.03.44

The finance domain also has much to gain from architectural cross-pollination with the tightly regulated health care sector. Kaissis et al. (2021) proposed PriMIA (Privacy-preserving Medical Image Analysis), an end-to-end open-source framework for secure FL with aggregated data. The configuration of virus-infective correlations in diseases as heterogeneous as gout is suggestive of specific virus-induced, injury-like pathologies. In like manner, Chen, Qin, et al. (2020) suggested FedHealth, a unique federated transfer learning framework. FedHealth serves as an ideal analogue for financial forecasting as a global model is allowed to be transmitted to edge devices, which is then fine-tuned through domain adaptation with local data. A generalized macroeconomic budget model in a financial context may be disseminated via FL and then adapted for specific organizational subsidiaries or regional branches, which is associated with a 5.3% accuracy gain over fixed global models (Chen, Qin, et al., 2020).

Table 3: Communication and Computational Optimization Strategies in Asynchronous FL

Optimization Strategy	**Theoretical Mechanism**	**Domain Application Context**	**Quantitative Efficiency Gain**
Triple-Step Asynchronous (TrisaFed)	Informative Client Activation (ICA) via Self-Relative Entropy (SRE)	High-volume transactional data streams	89.77% bandwidth reduction; >70% speedup
Layerwise Update & Aggregation	Deep network layers updated less frequently than shallow layers	Neural network forecasting models	Vast reduction in per-epoch parameter transmission
Adaptive Epoch Estimation	Computes local epochs dynamically based on specific client data volume	Resource-constrained edge agents	10.21s reduction in computational wait time per round

7. Conclusion

Creating financial budgets for multiple entities and forecasting their financial futures requires an architectural framework that does not depend on centralized data processing. After performing a comprehensive synthesis of the literature until 2022, it was found that Federated Learning when combined with strong cryptographic protocols and mathematical privacy frameworks offers a robust definite scalable solution.

10.48047/jocaaa.2023.31.03.44

Bayesian Differential Privacy and Multi-Key Homomorphic Encryption together tackle the tradeoff between privacy and utility. Through the use of Rényi divergence for data distributions and the distribution of decryption keys through the network, the two frameworks lower the privacy budget requirement to $\varepsilon<1$ and prevent parameter reconstruction while ensuring high algorithmic fidelity. Asynchronous strategies, for instance, TrisaFed, eliminate systemic limitations relating to network latency and communication overhead with more than 80% reduction in transmission costs.

Federated deep learning models outperform isolated legacy systems by more than 10% in sensitive and imbalanced applications (e.g. credit risk) with AUC metrics greater than 95.5% and an overall accuracy of greater than 99.9% in transactional fraud detection. Notably, game-theoretic structures such as MarS-FL align firms' incentives, which mathematically proves that direct rivals on the market can safely cooperate to obtain better predictive models without harming their commercial positions.

In time, these privacy-preserving artificial intelligence architectures will incorporate more and more RegTech and a bigger pool of secure open-source libraries like PySyft. Through the use of blockchain and the application of dynamic adversaries, financial institutions can embed a decentralized trust mechanism and reduce adversary vulnerabilities in order to ensure the global financial forecasting is resilient and transparent and according to law.

References

Abdel-Basset, M., Hawash, H., & Moustafa, N. (2022). Toward privacy preserving federated learning in Internet of Vehicular Things: Challenges and future directions. *IEEE Consumer Electronics Magazine*, *11*(6), 56–66. https://doi.org/10.1109/mce.2021.3117232

Alazab, M., RM, S. P., M, P., Maddikunta, P. K. R., Gadekallu, T. R., & Pham, Q.-V. (2022). Federated learning for cybersecurity: Concepts, challenges, and future directions. *IEEE Transactions on Industrial Informatics*, *18*(5), 3501–3509. https://doi.org/10.1109/tii.2021.3119038

Byrd, D., & Polychroniadou, A. (2020). Differentially private secure multi-party computation for federated learning in financial applications. *Proceedings of the First ACM International Conference on AI in Finance*, 1–9. https://doi.org/10.1145/3383455.3422562

10.48047/jocaaa.2023.31.03.44

Chen, Y., Qin, X., Wang, J., Yu, C., & Gao, W. (2020). FedHealth: A federated transfer learning framework for wearable healthcare. *IEEE Intelligent Systems*, *35*(4), 83–93. https://doi.org/10.1109/mis.2020.2988604

Chen, Y., Sun, X., & Jin, Y. (2020). Communication-efficient federated deep learning with layerwise asynchronous model update and temporally weighted aggregation. *IEEE Transactions on Neural Networks and Learning Systems*, *31*(10), 4229–4238. https://dacaoi.org/10.1109/tnnls.2019.2953131

Dasari, S. V., Mittal, K., GVK, S., Bapat, J., & Das, D. (2021). Privacy enhanced energy prediction in smart building using federated learning. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 1–6. https://doi.org/10.1109/iemtronics52119.2021.9422544

Imteaj, A., & Amini, M. H. (2022). Leveraging asynchronous federated learning to predict customers financial distress. *Intelligent Systems with Applications*, *14*, 200064. https://doi.org/10.1016/j.iswa.2022.200064

Kaissis, G., Ziller, A., Passerat-Palmbach, J., Ryffel, T., Usynin, D., Trask, A., Lima, I., Jr., Mancuso, J., Jungmann, F., Steinborn, M.-M., Saleh, A., Makowski, M., Rueckert, D., & Braren, R. (2021). End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence*, *3*(6), 473–484. https://doi.org/10.1038/s42256-021-00337-8

Liang, Y., Liu, Z., Song, Y., Yang, A., Ye, X., & Ouyang, Y. (2021). A methodology of trusted data sharing across telecom and finance sector under China's data security policy. *2021 IEEE International Conference on Big Data (Big Data)*, 5406–5412. https://doi.org/10.1109/bigdata52589.2021.9671857

Lim, W. Y. B., Luong, N. C., Hoang, D. T., Jiao, Y., Liang, Y., Yang, Q., Niyato, D., & Miao, C. (2020). Federated learning in mobile edge networks: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *22*(3), 2031–2063. https://doi.org/10.1109/comst.2020.2986024

Lv, B., Cheng, P., Zhang, C., Ye, H., Meng, X., & Wang, X. (2021). Research on modeling of e-banking fraud account identification based on federated learning. *2021 IEEE Intl Conf on Dependable, Autonomic and Secure Computing*

10.48047/jocaaa.2023.31.03.44

(DASC/PiCom/CBDCom/CyberSciTech), 611–618. https://doi.org/10.1109/dasc-picom-cbdcom-cyberscitech52372.2021.00105

Ma, J., Naas, S.-A., Sigg, S., & Lyu, X. (2022). Privacy-preserving federated learning based on multi-key homomorphic encryption. *International Journal of Intelligent Systems*, *37*(9), 5880–5901. https://doi.org/10.1002/int.22818

Myalil, D., Rajan, M. A., Apte, M., & Lodha, S. (2021). Robust collaborative fraudulent transaction detection using federated learning. *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 373–378. https://doi.org/10.1109/icmla52953.2021.00064

Nguyen, D. C., Ding, M., Pathirana, P. N., Seneviratne, A., Li, J., & Poor, H. V. (2021). Federated learning for Internet of Things: A comprehensive survey. *IEEE Communications Surveys & Tutorials*, *23*(3), 1622–1658. https://doi.org/10.1109/comst.2021.3075439

Shingi, G. (2020). A federated learning based approach for loan defaults prediction. *2020 International Conference on Data Mining Workshops (ICDMW)*, 362–368. https://doi.org/10.1109/icdmw51313.2020.00057

Triastcyn, A., & Faltings, B. (2019). Federated learning with Bayesian differential privacy. *2019 IEEE International Conference on Big Data (Big Data)*, 2587–2596. https://doi.org/10.1109/bigdata47090.2019.9005465

Wu, X., & Yu, H. (2022). MarS-FL: Enabling competitors to collaborate in federated learning. *IEEE Transactions on Big Data*, *10*(6), 801–811. https://doi.org/10.1109/tbdata.2022.3186991

Yang, F., Qiao, Y., Abedin, M. Z., & Huang, C. (2022). Privacy-preserved credit data sharing integrating blockchain and federated learning for industrial 4.0. *IEEE Transactions on Industrial Informatics*, *18*(12), 8755–8764. https://doi.org/10.1109/tii.2022.3151917

You, L., Liu, S., Chang, Y., & Yuen, C. (2022). A triple-step asynchronous federated learning mechanism for client activation, interaction optimization, and aggregation enhancement. *IEEE Internet of Things Journal*, *9*(23), 24199–24211. https://doi.org/10.1109/jiot.2022.3188556

10.48047/jocaaa.2023.31.03.44

Yu, H., Liu, Z., Liu, Y., Chen, T., Cong, M., Weng, X., Niyato, D., & Yang, Q. (2020). A sustainable incentive scheme for federated learning. *IEEE Intelligent Systems*, *35*(4), 58–69. https://doi.org/10.1109/mis.2020.2987774

Zhang, J., Chen, B., Cheng, X., Binh, H. T. T., & Yu, S. (2021). PoisonGAN: Generative poisoning attacks against federated learning in edge computing systems. *IEEE Internet of Things Journal*, *8*(5), 3310–3322. https://doi.org/10.1109/jiot.2020.3023126

Zhang, Y., Xie, Z., & Qu, X. (2022). Application and legal governance of computational technologies in credit investigation: An interdisciplinary perspective. *Discrete Dynamics in Nature and Society*, *2022*, 6278317. https://doi.org/10.1155/2022/6278317

Ziller, A., Trask, A., Lopardo, A., Szymkow, B., Wagner, B., Bluemke, E., Nounahon, J.-M., Passerat-Palmbach, J., Prakash, K., Rose, N., Ryffel, T., Reza, Z. N., & Kaissis, G. (2021). PySyft: A library for easy federated learning. In *Federated learning systems: Analysis, algorithms and applications* (pp. 111–139). Springer. https://doi.org/10.1007/978-3-030-70604-3_5