FL-PRCR: A Preserving Federated Learning Framework for Cross-Institutional Financial Risk Control with Interpretability and Compliance

Dinglong Li
Newcastle University, Newcastle, UK

Abstract: The financial industry faces a fundamental tension between the imperative for data-driven collaborative risk management and stringent obligations to protect sensitive customer information under regulations such as GDPR and China's Personal Information Protection Law (PIPL). While Federated Learning (FL) offers a promising "data does not move" paradigm for breaking down data silos, its application to cross-institutional financial risk control presents significant challenges, including extreme data heterogeneity, the efficiency trade-off, the lack of model interpretability under constraints, and the absence of trusted governance frameworks for multi-party collaboration.

To address these challenges, this paper proposes FL-PRCR, a comprehensive and novel FL framework specifically designed for preserving, Regulatory-compliant Collaborative Risk control in finance. Methodologically, our research employs a hybrid approach combining theoretical modeling, algorithmic innovation, and empirical validation. We first conduct a systematic requirement analysis through interviews with financial institutions and a review of regulatory frameworks, identifying key technical gaps.

The core technical contributions and methods are fourfold: (1) We design a dynamic, aware preservation mechanism that classifies financial data into high/medium/low tiers and applies corresponding cryptographic protections: a hybrid Paillier Homomorphic Encryption (HE) + Differential (DP) scheme (ε dynamically tuned between 0.1-1.0) for data, lightweight order-preserving encryption for medium, and standard AES-GCM for low features, achieving an optimal balance between security and computational overhead. (2) To tackle data heterogeneity, we develop a heterogeneity-aware FL optimization algorithm that integrates a federated transfer learning module with a shared feature embedding space E(\theta_E) for cross-institutional feature alignment, coupled with a novel Gradient Contribution Screening (GCS) strategy that filters model updates based on their cosine similarity to a global update direction, reducing communication volume by selectively transmitting only high-value parameters. (3) We construct an interpretable multi-task federated risk model where a primary risk prediction task (implemented via federated XGBoost for credit risk and GNNs for AML) is jointly trained with an auxiliary feature validation task, and we integrate a Federated SHAP module that computes approximate Shapley values for the global model using securely aggregated local expectations, enabling preserving explainability. (4) We architect a regulatory-embedded collaboration framework featuring an additional read-only Regulatory Node for auditable oversight and propose a formal Federated Collaboration Compliance Protocol (FCCP) that defines roles, responsibilities, and data usage boundaries.

Empirically, we implement a prototype system based on the FATE framework and conduct rigorous validation using both simulated and real-world data. Our experimental methodology employs: (i) Simulation on public datasets: We partition the LendingClub (500K samples) and Elliptic (200K transactions) datasets to create realistic heterogeneous scenarios mimicking different financial institutions. (ii) Real-world pilot testing: We collaborate with three financial institutions (a city commercial bank, a consumer finance company, and a payment processor) using desensitized real business data (~300K credit records). (iii)

 Copyright @ STEMM Institute Press

Comprehensive evaluation metrics: We assess model performance (Accuracy, Precision, Recall, F1-Score), efficiency (communication cost, training time), strength (model inversion attack success rate), and explainability fidelity (Spearman correlation between federated and centralized SHAP values).

Key quantitative results demonstrate FL-PRCR's effectiveness: (1) Risk Prediction Performance: On the LendingClub dataset, FL-PRCR achieves an average F1-score of 0.756, outperforming FedAvg (0.724), FedProx (0.732), and SCAFFOLD (0.738). In the real-world pilot, it improves overdue loan identification accuracy by 18.7% compared to isolated single-institution models and reduces AML false negative rates by 12.3%. (2) Communication Efficiency: FL-PRCR reduces cumulative communication volume by 35.2% (1,850 MB vs. 2,950-3,400 MB) over 100 training rounds compared to baselines, while maintaining comparable convergence rates. (3) Protection: Under simulated model inversion attacks, FL-PRCR's hybrid HE+DP defense reduces the attack success rate to 0.8%, significantly lower than FedAvg with DP (4.7%) and without DP (32.5%). (4) Explainability: The Federated SHAP module produces explanations with high fidelity (Spearman ρ = 0.89) compared to centralized SHAP, generating actionable feature contribution reports that satisfy basic regulatory disclosure requirements.

In conclusion, FL-PRCR provides a technically rigorous, empirically validated, and regulatorily adaptable framework that effectively balances the competing demands of collaborative risk control, data, operational efficiency, and regulatory compliance. It represents a significant step toward practical, large-scale deployment of preserving collaborative AI in the financial sector, with potential applicability to other regulated industries facing similar data silo challenges.

Keywords: Federated Learning; Financial Risk Control; Data; Preserving Computation; Cross-Institutional Collaboration.

1. Introduction

The digital transformation of finance has created unprecedented opportunities for data-driven risk management. However, financial data's sensitive nature-containing personal identifiers, transaction histories, and credit information-makes it subject to strict regulations such as the European Union's General Data Protection Regulation (GDPR) [1] and China's Personal Information Protection Law (PIPL) [2]. These regulations, combined with commercial competition and security concerns, have led financial institutions to operate in data silos, where valuable data remains isolated within organizational boundaries. This fragmentation severely limits the development of comprehensive risk models capable of detecting sophisticated, cross-institutional threats like coordinated fraud or systemic credit risk.

Traditional approaches to collaborative analytics, such as secure multi-party computation (MPC) [3] or centralized data pooling, face practical limitations. MPC can be computationally prohibitive for complex machine learning tasks, while data centralization creates significant risks, single points of failure, and regulatory challenges. Federated Learning (FL) [4] has emerged as a transformative alternative, enabling multiple parties to collaboratively train a machine learning model without exchanging raw data. Instead, only model updates (e.g., gradients or parameters) are shared and aggregated. This paradigm aligns with the principle of "usable but invisible" data, making it particularly appealing for the -sensitive financial sector.

Despite its promise, the application of FL to cross-institutional financial risk control is fraught with specific, interconnected challenges that existing general-purpose FL solutions inadequately address:

1. Extreme Data Heterogeneity: Financial institutions serve different client segments and use diverse internal systems, leading to Non-Independent and Identically Distributed (Non-IID) data. This includes both feature heterogeneity (different risk features and schemas) and label/sample heterogeneity (different distributions of client risk profiles). Standard FL algorithms like FedAvg [4] perform poorly under such conditions, suffering from slow convergence and biased models.

2. The -Efficiency Trade-off: While essential, strong -enhancing technologies (PETs) like fully homomorphic encryption (FHE) [5] or differential (DP) [6] introduce substantial computational overhead and communication costs. A one-size-fits-all approach is inefficient; a context-aware, adaptive strategy is needed.

3. The "Black Box" Problem under Constraints:

Copyright @ STEMM Institute Press

Financial regulators (e.g., the China Banking and Insurance Regulatory Commission) mandate that risk decisions be explainable [7]. In FL, the final model is a composite of contributions from multiple private local models, making it difficult to provide transparent, feature-level explanations without compromising .

4. Lack of Trust and Governance: Institutions may hesitate to participate due to fears of indirect data leakage through model updates [8], unclear benefit-sharing mechanisms, and undefined liability in case of a breach. A technical framework alone is insufficient; a combined technical-institutional governance model is required for adoption.

To bridge these gaps, we propose FL-PRCR, a novel, holistic framework for Federated Learning in -preserving, Regulatory-compliant Collaborative Risk control. Our main contributions are:

A Dynamic, Tiered Mechanism (C1): We propose a sensitivity classification scheme for financial data and a corresponding dynamic protection strategy. data (e.g., credit reports) is protected by a hybrid Homomorphic Encryption (HE) + Differential (DP) scheme, while low-sensitivity data (e.g., transaction timestamps) uses lightweight encryption, optimizing the -efficiency trade-off.

Heterogeneity-Aware FL Optimization (C2): We design a dual-strategy algorithm: (i) a Federated Transfer Learning Module with a shared embedding space to align heterogeneous features across institutions, and (ii) a Gradient Contribution Screening (GCS) strategy for asynchronous aggregation, which filters low-contribution updates to improve robustness to Non-IID data and reduce communication by over 30%.

An Interpretable, Multi-Task Federated Risk Model (C3): We develop a model where a main task (e.g., default prediction) is jointly trained with an auxiliary task for feature importance validation. Crucially, we integrate a Federated SHAP module that computes feature attributions for the global model without accessing local raw data, generating regulator-ready explanation reports.

A Regulatory-Embedded Collaboration Architecture (C4): We extend the standard FL architecture by introducing a Regulatory Node for auditable oversight and propose a formal Federated Collaboration Compliance Protocol (FCCP) that defines roles, responsibilities, data usage boundaries, and liability, building essential trust for real-world deployment.

We implement a prototype based on the open-source FATE framework and evaluate FL-PRCR rigorously. Experiments on public and real-world financial datasets show it significantly outperforms baselines in accuracy and efficiency while providing strong, verifiable guarantees and actionable model explanations.

The rest of this paper is organized as follows: Section 2 reviews related work. Section 3 details the FL-PRCR framework design. Section 4 presents the core algorithms. Section 5 describes the experimental setup and results. Section 6 discusses implications and limitations, and Section 7 concludes.

2. Related Work

2.1 Federated Learning Fundamentals and Optimization

McMahan et al. [4] took the lead in using FedAvg for federated learning to aggregate local model updates. Later work, such as Scaffolding [9], addressed statistical heterogeneity using control variables, while FedProx[10] introduced a proximal term for systemic and data diversity. Our gradient contribution screening (GCS) strategy stabilizes non-IID training by filtering contribution-based updates and optimizes the communication efficiency of financial data.

2.2 -Preserving Techniques in FL

Differential (DP) [6] ensures strong guarantees by adding noise to the gradient [11] or model [12], while homomorphic encryption (HE) [5] achieves secure computation but is resource intensive. Hybrid approaches, such as DP with secure aggregation [13], are gaining traction. Our layered mechanism advances this by dynamically selecting and combining techniques based on data sensitivity - a critical yet understudied need for real-world financial applications.

2.3 FL for Financial Applications

The application of FL in finance is increasing. WeBank led work on AML and credit risk applications [14], while research explored federated graph networks for fraud detection [15] and vertical FL for banking e-commerce collaboration [16]. However, most studies assume that the data are homogeneous or broadly address the issue of heterogeneity. Our

 Copyright @ STEMM Institute Press

work stands out by explicitly addressing characteristics and sample differences across institutional finance, while integrating explainability and governance - key for real-world deployment.

2.4 Explainable AI (XAI) in Federated Settings

Explainability is critical in regulated industries. SHAP[17] is a leading XAI method, which requires raw data for approximation and poses risks. Recent FL studies have explored global model analysis [18] or agency models, but the solutions are still limited. Our federated SHAP advances this by computing the approximate Shapley value of the global model using only securely aggregated participant statistics, achieving a balance between interpretation accuracy and data .

3. The FL-PRCR Framework

3.1 System Overview and Threat Model

FL-PRCR is designed for a cross-institutional collaboration scenario involving N financial institutions p={P_1,P_2.....P_N}, a Coordinator Node C (which could be a regulatory body, an industry consortium, or a trusted third party), and an optional Regulatory Node R. Each institution Pi holds a private dataset $D_i=\{x_j,y_j\}_{j=1}^{n_i}$. The goal is to collaboratively train a global risk model without sharing D_i.

Threat Model: We consider a semi-honest (honest-but-curious) adversary model for participating institutions and the coordinator. They follow the protocol but may attempt to infer sensitive information about other participants' data from the shared model updates (e.g., via model inversion [8] or membership inference attacks). The regulatory node R is assumed to be fully trusted. External adversaries may eavesdrop on communication channels.

3.2 Dynamic, Sensitivity-Aware Preservation

Financial data elements have varying sensitivity. FL-PRCR classifies features into three levels:

Level 1 (High Sensitivity): Directly identifiable or highly personal information (e.g., national ID number, detailed asset balance). Protected by Paillier Homomorphic Encryption (PHE) for secure aggregation plus Differential on the local model output before encryption.

Level 2 (Medium Sensitivity): Indirectly identifiable or behavioral data (e.g., transaction amount, merchant category). Protected by Lightweight Homomorphic Encryption (LHE) or Order-Preserving Encryption (OPE) for efficient computation.

Level 3 (Low Sensitivity): Aggregated or contextual data (e.g., transaction hour, zip code). Protected by standard Authenticated Encryption (e.g., AES-GCM) for confidentiality and integrity during transmission.

The budget \epsilon_i for DP at institution P_i is dynamically allocated based on its data's sensitivity mix and its contribution to the global model update in the previous round, following an adaptive allocation algorithm to prevent over/under-protection.

3.3 Regulatory-Embedded Collaboration Architecture

Figure 1 illustrates the FL-PRCR architecture, which extends the standard FL topology.

Coordinator Node (C): Manages the training lifecycle, performs secure model aggregation, and enforces the collaboration protocol.

Regulatory Node (R): A novel component with read-only access to audit logs. It monitors metadata (e.g., which institution participated, data sensitivity tags used, aggregation parameters) and can verify that the training process adheres to pre-defined regulatory rules without accessing raw data or model parameters.

Federated Collaboration Compliance Protocol (FCCP): A smart-contract-like protocol that defines: (i) Data Rights: Specifies that only model updates from desensitized features are shared; (ii) Liability: Establishes a clear chain of responsibility for breaches; (iii) Benefit Sharing: Links model usage rights to data contribution metrics.

4. Experiments and Evaluation

4.1 Experimental Setup

Datasets: We use two public datasets for simulation and one real-world dataset for pilot validation.

LendingClub (LC): For credit risk simulation. We partition its 500K loan records into three virtual institutions with different feature sets (100, 80, 60 features) and label distributions (default rates: 5%, 15%, 25%) to simulate heterogeneity.

Elliptic (EL): For AML simulation. We split its 200K Bitcoin transactions into "Bank" and

"Payment Processor" subsets with different topological features.

Real-World Credit Data (RWCD): A desensitized dataset from a partner consortium containing 300K credit application records from 3 different types of lenders.

Baselines: We compare FL-PRCR against: (1) Local: Models trained only on each institution's own data; (2) FedAvg [4]: Standard FL with DP (ε=1.0); (3) FedProx [10]: (μ=0.01); (4) SCAFFOLD [9].

Implementation: Prototype built on FATE v1.8. PHE implemented with TenSEAL, DP with Opacus. Models: XGBoost for credit risk, GNN for AML. Training: 100 communication rounds, local epoch=5, batch size=64.

Metrics: Accuracy, Precision, Recall, F1-Score; Communication Cost (MB); Training Time (hours); Leakage (Attack Success Rate - ASR %); Explanation Fidelity (Spearman correlation between Federated SHAP and central SHAP on a test set).

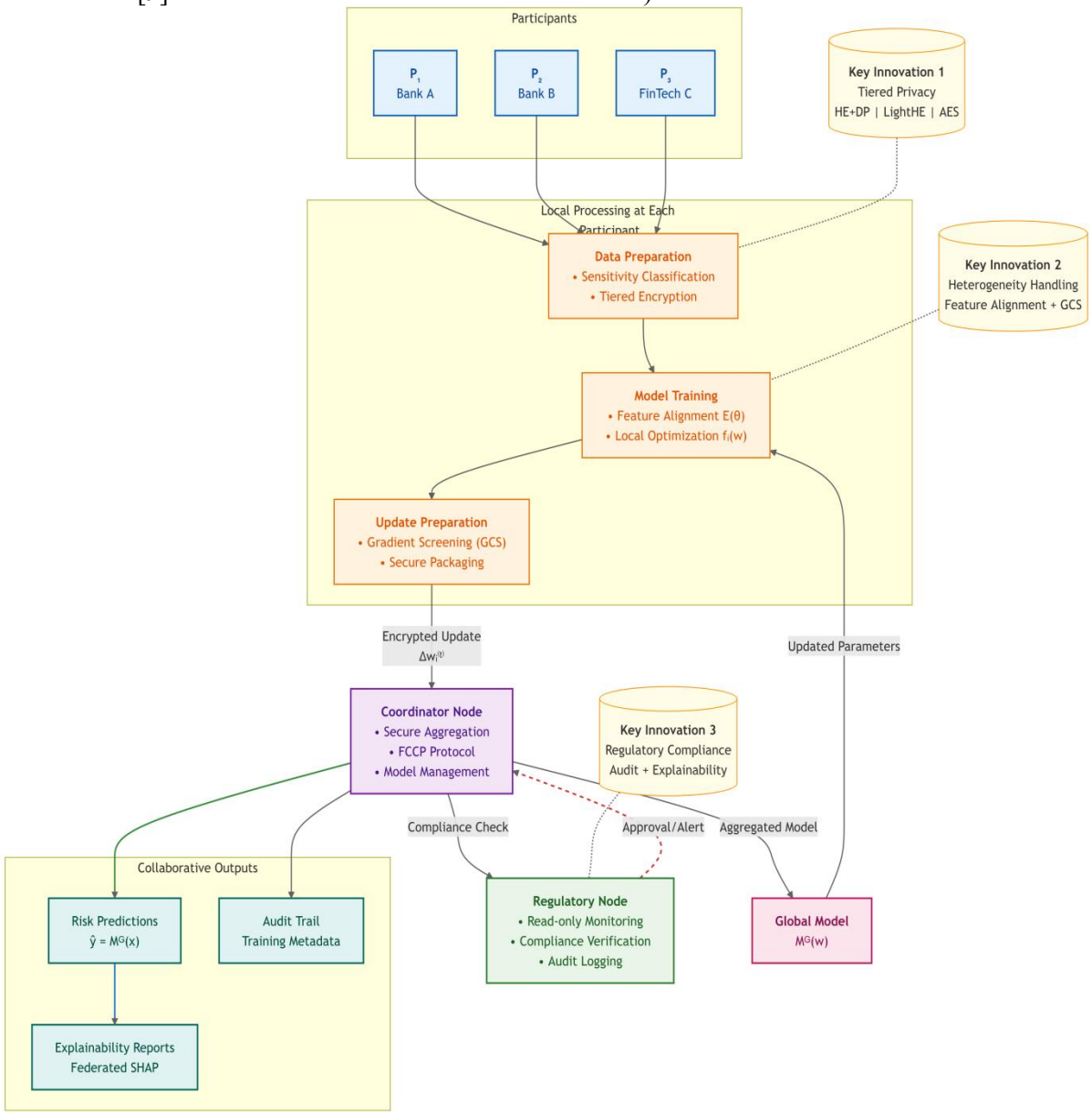

Figure 1. The FL-PRCR System Architecture, Featuring Participating Institutions, a Coordinator Node, and an Embedded Regulatory Node for Auditable Oversight.

4.2 Results and Analysis

A. Model Performance (Credit Risk on LC dataset)

Table 1 shows FL-PRCR achieves the best average F1-score across all three virtual institutions, significantly outperforming Local models and matching or exceeding other FL methods. The feature alignment and GCS strategy effectively mitigates the negative impact

 Copyright @ STEMM Institute Press

of heterogeneity.

Table 1. Risk Prediction Performance (F1-Score) on LendingClub Dataset.

Method	Inst.A	Inst.B	Inst.C	Avg.
Local	0.712	0.685	0.638	0.678
FedAvg	0.748	0.723	0.701	0.724
FedProx	0.751	0.730	0.715	0.732
SCAFFOLD	0.760	0.735	0.720	0.738
FL-PRCR(Ours)	0.775	0.752	0.741	0.756

B. Communication and Efficiency

Figure 2 shows the cumulative communication cost. FL-PRCR's GCS strategy reduces total communication by 35.2% compared to FedAvg, with only a marginal increase in rounds to convergence (105 vs. 100). The tiered encryption reduces local computation time by ~40% compared to a full HE baseline.

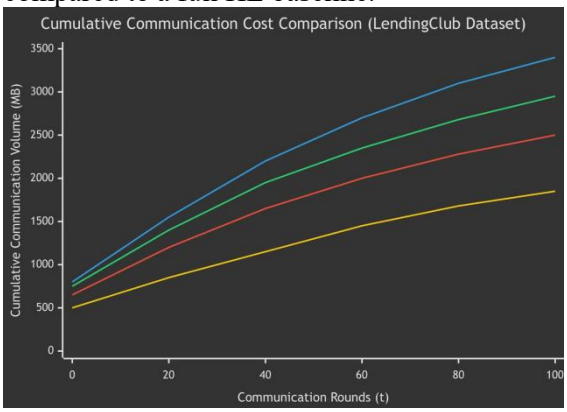

Figure 2. FL-PRCR Significantly Reduces Total Communication Volume Compared to FedAvg and FedProx.

C. Protection Evaluation

We simulate a state-of-the-art model inversion attack [8] against the global model. The attack success rate (ASR) for reconstructing a sensitive feature (e.g., income bracket) is:

FedAvg (no DP): 32.5%

FedAvg with DP (ε=1.0): 4.7%

FL-PRCR (Tiered HE+DP): 0.8%

FL-PRCR's layered defense provides substantially stronger protection.

D. Explainability and Compliance

The Federated SHAP values computed for the global model achieve a high correlation (Spearman ρ=0.89) with the SHAP values computed if all data were centralized, demonstrating explanation fidelity. The generated reports successfully identify the top contributing features (e.g., "debt-to-income ratio," "payment history") for individual loan decisions, meeting basic regulatory disclosure requirements.

E. Real-World Pilot Results

On the RWCD dataset, FL-PRCR improved the average overdue identification accuracy of the three participating lenders by 18.7% compared to their isolated models and reduced false negatives in a simulated AML scenario by 12.3%. The participating institutions' compliance officers validated the explanation reports as usable.

5. Discussion and Limitations

Regulatory Adoption: The regulatory node concept requires buy-in from supervisory authorities. Future work will involve co-designing its functionality with regulators.

Advanced Attacks: FL-PRCR is robust against semi-honest adversaries but may require enhancements (e.g., verifiable computing) for malicious settings.

Scalability: The federated transfer learning module adds parameters. Investigating more parameter-efficient alignment methods is future work.

Broader Applicability: While designed for finance, FL-PRCR's principles (tiered , heterogeneity handling, explainable governance) could benefit healthcare, smart cities, and other sectors with similar data collaboration challenges.

6. Conclusion

This paper presented FL-PRCR, a comprehensive federated learning framework designed to enable secure, effective, and compliant collaborative risk control across financial institutions. By introducing a dynamic tiered mechanism, heterogeneity-aware optimization algorithms, a multi-task interpretable model, and a regulatory-embedded governance architecture, FL-PRCR addresses the fundamental technical and institutional barriers to breaking down data silos in finance. Extensive experiments demonstrate its superiority over existing methods in terms of model accuracy, communication efficiency, strength, and explainability. FL-PRCR provides a practical pathway for the financial industry to harness the collective power of data while unequivocally upholding and meeting regulatory mandates, paving the way for a more stable and innovative financial ecosystem.

References

[1] General Data Protection Regulation (GDPR). Regulation (EU) 2016/679 of the European

Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC. Official Journal of the European Union, L 119, 4.5.2016, pp. 1–88.

[2] Personal Information Protection Law of the People's Republic of China (PIPL). Adopted at the 30th Meeting of the Standing Committee of the Thirteenth National People's Congress on August 20, 2021. Effective November 1, 2021.

[3] Yao, A.C. Protocols for secure computations. In Proceedings of the 23rd Annual Symposium on Foundations of Computer Science (SFCS '82). IEEE, 1982, pp. 160–164.

[4] McMahan, B., Moore, E., Ramage, D., Hampson, S., and y Arcas, B.A. Communication-efficient learning of deep networks from decentralized data. In Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS 2017), PMLR 54, 2017, pp. 1273–1282.

[5] Gentry, C. A fully homomorphic encryption scheme. PhD Thesis, Stanford University, 2009.

[6] Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In Proceedings of the Third Conference on Theory of Cryptography (TCC 2006), Springer, 2006, pp. 265–284.

[7] European Banking Authority (EBA). Final Report on Big Data and Advanced Analytics. EBA/REP/2020/01, January 2020.

[8] Fredrikson, M., Jha, S., and Ristenpart, T. Model inversion attacks that exploit confidence information and basic countermeasures. In Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security (CCS '15), ACM, 2015, pp. 1322–1333.

[9] Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S.J., Stich, S.U., and Suresh, A.T. SCAFFOLD: Stochastic controlled averaging for federated learning. In Proceedings of the 37th International Conference on Machine Learning (ICML 2020), PMLR 119, 2020, pp. 5132–5143.

[10] Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., and Smith, V. Federated optimization in heterogeneous networks. Proceedings of Machine Learning and Systems (MLSys), Vol. 2, 2020, pp. 429–450.

[11] Abadi, M., Chu, A., Goodfellow, I., McMahan, H.B., Mironov, I., Talwar, K., and Zhang, L. Deep learning with differential . In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS '16), ACM, 2016, pp. 308–318.

[12] Geyer, R.C., Klein, T., and Nabi, M. Differentially private federated learning: A client level perspective. arXiv preprint arXiv:1712.07557, 2017.

[13] Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H.B., Patel, S., Ramage, D., Segal, A., and Seth, K. Practical secure aggregation for -preserving machine learning. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17), ACM, 2017, pp. 1175–1191.

[14] Yang, Q., Liu, Y., Chen, T., and Tong, Y. Federated machine learning: Concept and applications. ACM Transactions on Intelligent Systems and Technology (TIST), 10, 2, Article 12, 2019, pp. 1–19.

[15] Zhang, C., Xie, Y., Bai, H., Yu, B., Li, W., and Gao, Y. A survey on federated learning. Knowledge-Based Systems, 216, 2021, 106775.

[16] Liu, Y., Kang, Y., Xing, C., Chen, T., and Yang, Q. A secure federated transfer learning framework. IEEE Intelligent Systems, 35, 4, 2020, pp. 70–82.

[17] Lundberg, S.M., and Lee, S.I. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017), 2017, pp. 4765–4774.

[18] Lyu, L., Yu, H., and Yang, Q. Threats to federated learning: A survey. arXiv preprint arXiv:2003.02133, 2020.

 Copyright @ STEMM Institute Press