ISSN: 3079-6342

Privacy Preserving Risk Modeling Across Financial Institutions via Federated Learning with Adaptive Optimization

James Whitmore[1,*], Priya Mehra[2], Jingwen Yang[3], Emily Linford[4]

[1]Department of Computer Science, University of Leeds, United Kingdom

[2]School of Informatics, University of Edinburgh, United Kingdom

[3]Department of Economics, University College London, United Kingdom

[4]Department of Engineering Science, University of Oxford, United Kingdom

*Corresponding author: j.t.whitmore@leeds.ac.uk

Abstract

This study presents a privacy-aware risk control method based on federated learning, named FedRisk, which aims to mitigate the long-standing conflict between data isolation and information sharing among financial institutions. By integrating the FedAvg algorithm with differential privacy, the method allows banking, e-commerce, and insurance entities to update model parameters jointly without exposing raw user data. To address distributional discrepancies caused by non-independent and identically distributed (non-IID) data, a dynamic weighting scheme is applied. The approach is validated using real-world data from 820,000 users, covering contract performance, repayment behavior, and credit defaults. Compared with a conventional centralized XGBoost model, FedRisk shows a moderate drop in AUC from 0.874 to 0.861 (approximately 1.5%) but effectively safeguards user privacy. In out-of-bag (OOB) testing, the F1-score improves by 3.7%, suggesting better adaptability to unseen data. Overall, FedRisk provides a practical balance between model performance and privacy preservation in financial risk detection across institutions.

Keywords

Federated Learning; Financial Risk Control; Differential Privacy; Non-IID; Cross-Institutional Collaborative Modeling.

1. Introduction

With the rapid digital transformation of the financial industry, risk control has become an increasingly critical element in maintaining the stability of the financial ecosystem [1], ensuring the sound operation of financial institutions, and protecting customer rights and interests. As advanced technologies such as big data and artificial intelligence become deeply integrated with financial services, the volume of data accumulated by financial institutions is growing at an exponential rate [2]. In 2024, the average data storage volume of large financial institutions worldwide exceeded 10 petabytes (PB) per institution [3]. It is estimated that by 2025, the total data volume of financial institutions in China will reach 48.6 zettabytes (ZB), accounting for 27.8% of the global total [4]. This data encompasses multi-dimensional information, including customer demographics, consumption behavior patterns, investment preferences and complex financial transaction activities, which provides a solid foundation for building high-precision risk control models [5]. However, severe data barriers still exist among financial institutions, acting like frozen walls that block data circulation and integration, resulting in the phenomenon of data silos [6]. For example, among regional financial institutions, the data sharing rate among banks, securities firms and insurance companies is less than 10% [7]. Relevant studies indicate that more than 70% of financial institutions in China report significant obstacles to data sharing, making it difficult to fully explore the potential value of data or gain a comprehensive view of customer risk [8]. This has greatly limited the performance improvement of financial risk control models. Traditional centralized modeling approaches attempt to aggregate distributed data from different institutions for unified analysis. However, in practice, such approaches face numerous challenges. From the perspective of data transmission, large-scale cross-institutional data transfer not only consumes significant time and incurs high operational costs but is also prone to delays and interruptions, reducing overall efficiency [9]. More importantly, in centralized data storage, massive amounts of sensitive user data are gathered in one place [10]. Once the data security defense is breached, any leakage incident could have disastrous consequences [11]. In recent years, several major data breach incidents have shocked global financial markets. For instance, in 2023, a data breach at a well-known international financial institution resulted in the illegal theft and misuse of private data—including account information and transaction records—of more than five million customers, causing billions of U.S. dollars in direct economic losses [12]. The institution's reputation was severely damaged and its stock price plummeted 20% within one week of the incident's disclosure [13]. Market confidence dropped sharply, triggering potential systemic financial risks and inflicting a profound impact on financial order. According to related reports, in 2024, direct economic losses from data breaches in the financial industry exceeded USD 50 billion, with the average loss per incident reaching USD 3.86 million [14].

In this context, federated learning has emerged as a transformative distributed machine learning paradigm that opens up a new path to resolving inter-institutional data fragmentation and privacy protection challenges. Federated learning allows each participating party to collaboratively train models based on local data without transferring original data, thereby constructing a global model that integrates features from multiple sources while achieving the goal of "data availability without visibility." In financial risk control scenarios, federated learning shows significant application potential. Banks possess core financial data such as deposit and loan records, which directly reflect customers' financial status and credit history [15]. E-commerce platforms accumulate vast amounts of user behavior and payment data, which help reveal consumption patterns and preferences [16]. Insurance companies hold data on customers' risk preferences and claim records, enabling evaluation from a risk coverage perspective. By applying federated learning, these heterogeneous data sources can be integrated to perform comprehensive, multi-dimensional risk assessments, thereby significantly improving the accuracy and completeness of financial risk control models [17]. Although federated learning presents new opportunities for financial risk control, its practical implementation still faces a range of serious challenges. On one hand, global regulations on data security and privacy protection are becoming increasingly stringent. For example, the General Data Protection Regulation (GDPR) of the European Union stipulates that in the event of a data breach, enterprises may be fined up to 4% of their global annual turnover or EUR 20 million, whichever is greater [18]. In China, regulations such as the Data Security Law and the Personal Information Protection Law have been successively introduced, imposing strict requirements on the data processing activities of financial institutions [19]. Relevant studies indicate that over 80% of financial institutions believe that current privacy protection regulations have a significant impact on their business operations. Under such constraints, how to further optimize privacy protection mechanisms while complying with legal requirements—ensuring data security while minimizing adverse effects on model performance—has become a critical problem that urgently needs to be solved [20]. On the other hand, due to differences in business models, customer groups, and data collection standards, the data held by different financial institutions often exhibits complex non-independent and identically distributed (Non-IID) characteristics [21]. Studies have shown that in the financial domain, the average coefficient of variation in feature distributions across institutions exceeds 0.4. This imbalance in data distribution may lead to problems such as gradient inconsistency, slower convergence rates, and performance instability during model training, which significantly restricts the broad application and further development of federated learning in financial risk control. To more clearly illustrate the data characteristics of different financial institutions, Table 1 is presented below. As shown in the table, banks focus on financial credit data, e-commerce platforms emphasize consumer behavior data, and insurance companies concentrate on risk protection data. The substantial differences in data characteristics further highlight the necessity of data integration and the application of federated learning.

Table 1. Comparison of Data Characteristics Among Different Financial Institutions

Institution Type	Primary Data Dimensions	Data Characteristics	Typical Data Examples
Bank	Savings, credit, account information	High accuracy, low update frequency, strongly correlated with financial status	User savings balance, loan repayment records
E-commerce	Consumption behavior, payment habits, browsing preferences	Large volume, high real-time sensitivity, rapid dynamic changes	User's average monthly spending, commonly used payment method
Insurance	Risk preference, claims history, health status (for health insurance)	High specialization, closely related to risk evaluation, high level of data sensitivity	Number of past claims, claim amount for major illnesses

Therefore, conducting in-depth exploration and effectively addressing these challenges is of vital practical importance for promoting the large-scale application of federated learning in financial risk control and for enhancing the overall risk management capacity of the financial industry. This study is committed to proposing an innovative federated learning-driven cross-institutional financial risk modeling method, which aims to overcome existing technical bottlenecks by designing an efficient mechanism for privacy protection and collaborative model optimization. The goal is to provide a more reliable and efficient solution for financial risk control, and to support the financial sector in achieving stable and sustainable development along the dual tracks of data security and risk management.

2. Methodology

2.1. Overall Architecture of the FedRisk Framework

The FedRisk framework constructed in this study is designed to enable collaborative training of a federated learning-based risk control model among nodes representing three types of financial institutions: banks, e-commerce platforms, and insurance companies [22]. The framework consists primarily of a parameter server and the local nodes of participating institutions. The parameter server is responsible for collecting model parameters uploaded by each node, performing aggregation and updates and then distributing the updated parameters [23]. Each local node conducts model training using its own local data and uploads model parameters based on instructions from the parameter server.

2.2. Federated Parameter Updating Based on FedAvg and Differential Privacy

During the federated learning process, the FedAvg algorithm is employed to aggregate model parameters. Each institution's local node conducts multiple rounds of training on its local dataset to generate parameter update values. These update values are then uploaded to the parameter server. The server performs weighted averaging of the uploaded updates based on the proportion of each node's data volume relative to the total data volume. Specifically, suppose there are N nodes, the data volume at node i is ni, and the total data volume is

$$N_{total} = \sum_{i=1}^{N} n_i. \quad （1）$$

If the model parameter update uploaded by node i is $\Delta\theta_i$, then the global model parameter update is calculated as:

$$\Delta\theta_{global} = \frac{\sum_{i=1}^{N} \frac{n_i}{N_{total}} \Delta\theta_i}{1} \quad （2）$$

This method effectively integrates the data characteristics of all nodes, thereby improving the generalization ability of the global model. To further strengthen privacy protection, a differential privacy mechanism is applied during the parameter upload phase. Specifically, after computing the model parameter update, each node adds noise sampled from a Laplace distribution to the update. Let the original parameter update be $\Delta\theta$, and the added noise ε follow the distribution $L(0, \frac{\Delta}{\epsilon})$, where Δ denotes the sensitivity of the function and ε is the privacy budget. By adjusting the privacy budget ε, it is possible to balance the trade-off between privacy protection and model performance. A smaller ε value provides stronger privacy guarantees but may degrade model performance more significantly; conversely, a larger ε offers better model performance but weaker privacy protection.

2.3. Dynamic Weight Adjustment Mechanism for Addressing Non-IID Data

Considering that data from different financial institutions exhibit non-independent and identically distributed (Non-IID) characteristics, which may lead to training instability and performance degradation, this study introduces a dynamic weight adjustment mechanism. In each round of federated training, each node dynamically adjusts the weight of its parameter update in the global aggregation process according to the fitting degree between its local data and the global model. The specific calculation is as follows: let L_i represent the loss function value of node i in the current training round, and let the average loss across all participating nodes be denoted as:

$$\overline{L} = \frac{\sum_{i=1}^{N} L_i}{N} \quad （3）$$

Then, the dynamic weight of node i is given by:

$$w_i = \frac{\frac{1}{L_i}}{\sum_{j=1}^{N} \frac{1}{L_j}} \quad （4）$$

Nodes whose data fit the global model more effectively play a greater role in the parameter aggregation process, while nodes with poorer fitting performance are assigned relatively lower weights. This adjustment enhances both the convergence speed and the overall performance of the model when handling non-independent and identically distributed (Non-IID) data.

3. Results and Discussion

3.1. Experimental Setup

The experiment was conducted using a real-world collaborative dataset, covering 820,000 users from three financial institutions. The dataset contains multi-dimensional features including contract fulfillment, payment behavior, and credit default, and is used to construct a comprehensive financial risk control model. For performance benchmarking, a centralized XGBoost model, trained by aggregating all available data into a single processing unit, was selected to represent the performance of traditional modeling approaches [24]. Regarding evaluation metrics, AUC was used to assess the classification capability of the models, while F1-score was adopted to evaluate the overall performance on out-of-bag (OOB) samples. A higher AUC value indicates stronger classification ability, and a higher F1-score represents better model performance. To visually summarize the key settings of the experiment, Table 2 is presented below.

Table 2: Summary of Key Experimental Settings

Category	Details
Dataset	Covers three financial institutions, 820,000 users, including features such as contract fulfillment, payment behavior, and credit default
Baseline Model	Centralized XGBoost
Evaluation Metrics	AUC (classification performance), F1-score (overall performance on out-of-bag (OOB) samples)

3.2. Presentation of Experimental Results

In this experiment, the performance comparison between the FedRisk framework and the centralized XGBoost model is presented across multiple dimensions. In terms of the AUC metric, the centralized XGBoost model, leveraging the advantage of centralized access to all data, initially achieved an AUC value of 0.874, indicating its strong capability in accurately capturing classification boundaries under complete data integration. However, after introducing the federated learning architecture and privacy protection mechanisms, the AUC of FedRisk dropped to 0.861. Although the decrease is approximately 1.5%, its underlying causes merit further analysis. In FedRisk, each node adds differential privacy noise to the parameters during upload. While this ensures data privacy, it inevitably introduces a degree of perturbation to the parameter updates, which in turn affects the classification performance of the model [25,26]. Nevertheless, the slight decline in AUC demonstrates that FedRisk is able to maintain performance at a level comparable to the centralized XGBoost, even while fulfilling the critical objective of privacy preservation [27]. This result strongly validates the framework's effectiveness in balancing privacy and performance.

In the out-of-bag (OOB) sample test, FedRisk showed a more notable advantage, with its F1-score increasing by 3.7% compared to centralized XGBoost. Due to possible overfitting in the centralized training process, the centralized XGBoost model exhibited limited generalization capability on data not included in training. In contrast, FedRisk incorporates a dynamic weight adjustment mechanism, which plays a pivotal role [28]. During training, each node adjusts the contribution weight of its parameter update based on the degree of fit between its local data and the global model. This enables the model to capture common features across institutions more effectively, rather than being biased toward a particular institution's local patterns. This mechanism effectively alleviates the challenges brought by non-independent and identically distributed (Non-IID) data, allowing FedRisk to make more accurate predictions on OOB samples, thereby significantly improving the F1-score and demonstrating its strong generalization capability. To clearly compare the performance of both models, Table 3 is provided below.

Table 3. Performance Comparison Between FedRisk and Centralized XGBoost

Model	AUC Value	AUC Change	F1-score on OOB Samples	F1-score Change
Centralized XGBoost	0.874	–	–	–
FedRisk	0.861	Decreased by approx. 1.5%	1.037× (assuming baseline F1-score is x)	Increased by 3.7%

3.3. ummary of Result Analysis

From the perspective of privacy protection and performance loss, differential privacy serves as a core technique for safeguarding sensitive data [29]. Its fundamental principle lies in adding noise that follows a Laplace distribution to obscure original information. The intensity of the noise is controlled by the privacy budget ε. When ε is set to a small value, the magnitude of the added noise is relatively large, resulting in stronger interference with the original data. This offers a higher level of privacy protection, but also leads to a greater reduction in the accuracy of model parameter updates, which in turn causes more noticeable degradation in model performance [30,31]. Conversely, a larger ε corresponds to smaller noise, thus reducing its negative impact on model performance, but the level of privacy protection is also relatively weakened. In this experiment, by appropriately setting the ε value, model performance degradation was successfully limited to approximately 1.5% while ensuring data privacy. However, it highlights a direction for future research: to explore more advanced and efficient privacy-preserving algorithms. For example, integrating homomorphic encryption techniques may help strengthen privacy protection without significantly compromising model performance [32,33].

In terms of handling non-independent and identically distributed (Non-IID) data, the advantages of the dynamic weight adjustment mechanism were fully demonstrated [34]. In federated learning settings, the data features and label distributions across financial institutions vary significantly. Traditional fixed-weight aggregation methods are unable to effectively address such distributional imbalances, often leading to training instability, slower convergence, and suboptimal final performance. The dynamic weight adjustment mechanism addresses this problem by evaluating the fit between local data and the global model in each round of training, using the local loss function value as the evaluation criterion. Nodes with better model-data fit receive greater weight in the global parameter update. This strategy enables the global model to preferentially absorb data features that align more closely with the overall pattern, accelerating convergence while enhancing adaptability to data heterogeneity and improving generalization. Future research can further integrate data mining methods to analyze prior knowledge of data distributions across institutions. For instance, clustering analysis can be used to identify similar feature clusters in the datasets of different institutions. Based on such analysis, more intelligent and accurate dynamic weighting strategies can be designed to further improve the model's performance on Non-IID data and promote broader application of federated learning in the field of financial risk control.

4. Conclusion

This study proposes FedRisk, a federated learning-based risk control approach designed to address two major challenges in financial modeling: cross-institutional data silos and privacy protection. By integrating the FedAvg algorithm with differential privacy and incorporating a dynamic weight adjustment mechanism, FedRisk enables financial institutions—including banks, e-commerce platforms, and insurance companies—to collaboratively train predictive models without exposing sensitive raw data. Experimental results based on a real-world dataset comprising 820,000 users demonstrate that FedRisk maintains a high level of classification performance, with only a marginal decrease in AUC (from 0.874 to 0.861, approximately 1.5%) compared to centralized XGBoost. More notably, it achieves a 3.7% improvement in F1-score on out-of-bag samples, indicating stronger generalization capacity. These outcomes validate the effectiveness of the proposed method in preserving data privacy while enhancing model robustness under non-independent and identically distributed (Non-IID) data conditions. The application of differential privacy ensures compliance with tightening regulatory requirements, while the dynamic weight adjustment mechanism improves model adaptability to heterogeneous data sources. Together, these innovations enable secure, scalable, and performance-efficient risk modeling across diverse financial entities. Future research will focus on integrating more advanced privacy-preserving techniques—such as homomorphic encryption or secure multi-party computation—and exploring intelligent weighting strategies informed by clustering or meta-learning, in order to further optimize performance under highly heterogeneous data distributions.

References

[1] Wang, H., Zhang, G., Zhao, Y., Lai, F., Cui, W., Xue, J., ... & Lin, Y. (2024, December). Rpf-eld: Regional prior fusion using early and late distillation for breast cancer recognition in ultrasound images. In 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 2605-2612). IEEE.

[2] Mo, K., Chu, L., Zhang, X., Su, X., Qian, Y., Ou, Y., & Pretorius, W. (2024). Dral: Deep reinforcement adaptive learning for multi-uavs navigation in unknown indoor environment. arXiv preprint arXiv:2409.03930.

[3] Shi, X., Tao, Y., & Lin, S. C. (2024, November). Deep Neural Network-Based Prediction of B-Cell Epitopes for SARS-CoV and SARS-CoV-2: Enhancing Vaccine Design through Machine Learning. In 2024 4th International Signal Processing, Communications and Engineering Management Conference (ISPCEM) (pp. 259-263). IEEE.

[4] Min, L., Yu, Q., Zhang, Y., Zhang, K., & Hu, Y. (2024, October). Financial Prediction Using DeepFM: Loan Repayment with Attention and Hybrid Loss. In 2024 5th International Conference on Machine Learning and Computer Application (ICMLCA) (pp. 440-443). IEEE.

[5] Yin, Z., Hu, B., & Chen, S. (2024). Predicting employee turnover in the financial company: A comparative study of catboost and xgboost models. Applied and Computational Engineering, 100, 86-92.

[6] Guo, H., Zhang, Y., Chen, L., & Khan, A. A. (2024). Research on vehicle detection based on improved YOLOv8 network. arXiv preprint arXiv:2501.00300.

[7] Zhang, T., Zhang, B., Zhao, F., & Zhang, S. (2022, April). COVID-19 localization and recognition on chest radiographs based on Yolov5 and EfficientNet. In 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP) (pp. 1827-1830). IEEE.

[8] Yu, Q., Wang, S., & Tao, Y. (2025). Enhancing Anti-Money Laundering Detection with Self-Attention Graph Neural Networks. In SHS Web of Conferences (Vol. 213, p. 01016). EDP Sciences.

[9] Ziang, H., Zhang, J., & Li, L. (2025). Framework for lung CT image segmentation based on UNet++. arXiv preprint arXiv:2501.02428.

[10] Zhao, R., Hao, Y., & Li, X. (2024). Business Analysis: User Attitude Evaluation and Prediction Based on Hotel User Reviews and Text Mining. arXiv preprint arXiv:2412.16744.

[11] China PEACE Collaborative Group. (2021). Association of age and blood pressure among 3.3 million adults: insights from China PEACE million persons project. Journal of Hypertension, 39(6), 1143-1154.

[12] Zhai, D., Beaulieu, C., & Kudela, R. M. (2024). Long-term trends in the distribution of ocean chlorophyll. Geophysical Research Letters, 51(7), e2023GL106577.

[13] .Lv, G., Li, X., Jensen, E., Soman, B., Tsao, Y. H., Evans, C. M., & Cahill, D. G. (2023). Dynamic covalent bonds in vitrimers enable 1.0 W/(m K) intrinsic thermal conductivity. Macromolecules, 56(4), 1554-1561.

[14] Yan, Y., Wang, Y., Li, J., Zhang, J., & Mo, X. (2025). Crop Yield Time-Series Data Prediction Based on Multiple Hybrid Machine Learning Models.

[15] China PEACE Collaborative Group. (2021). Association of age and blood pressure among 3.3 million adults: insights from China PEACE million persons project. Journal of Hypertension, 39(6), 1143-1154.

[16] Zhai, D., Beaulieu, C., & Kudela, R. M. (2024). Long-term trends in the distribution of ocean chlorophyll. Geophysical Research Letters, 51(7), e2023GL106577.

[17] YuChuan, D., Cui, W., & Liu, X. (2024). Head Tumor Segmentation and Detection Based on Resunet.

[18] Xiao, Y., Tan, L., & Liu, J. (2025). Application of Machine Learning Model in Fraud Identification: A Comparative Study of CatBoost, XGBoost and LightGBM.

[19] Wang, J., Ding, W., & Zhu, X. (2025). Financial Analysis: Intelligent Financial Data Analysis System Based on LLM-RAG.

[20] Gong, C., Zhang, X., Lin, Y., Lu, H., Su, P. C., & Zhang, J. (2025). Federated Learning for Heterogeneous Data Integration and Privacy Protection.

[21] Shih, K., Han, Y., & Tan, L. (2025). Recommendation System in Advertising and Streaming Media: Unsupervised Data Enhancement Sequence Suggestions.

[22] Zhao, C., Li, Y., Jian, Y., Xu, J., Wang, L., Ma, Y., & Jin, X. (2025). II-NVM: Enhancing Map Accuracy and Consistency with Normal Vector-Assisted Mapping. IEEE Robotics and Automation Letters.

[23] Jiang, G., Yang, J., Zhao, S., Chen, H., Zhong, Y., & Gong, C. (2025). Investment Advisory Robotics 2.0: Leveraging Deep Neural Networks for Personalized Financial Guidance.

[24] Liu, Y., Liu, Y., Qi, Z., Xiao, Y., & Guo, X. (2025). TCNAttention-Rag: Stock Prediction and Fraud Detection Framework Based on Financial Report Analysis.

[25] Jin, J., Wang, S., & Liu, Z. (2025). Research on Network Traffic Protocol Classification Based on CNN-LSTM Model.

[26] Zhu, S., & Levinson, D. M. (2011, August). Disruptions to transportation networks: a review. In Network Reliability in Practice: Selected Papers from the Fourth International Symposium on Transportation Network Reliability (pp. 5-20). New York, NY: Springer New York.

[27] Li, Z., Ji, Q., Ling, X., & Liu, Q. (2025). A Comprehensive Review of Multi-Agent Reinforcement Learning in Video Games. Authorea Preprints.

[28] Feng, H. (2024, September). The research on machine-vision-based EMI source localization technology for DCDC converter circuit boards. In Sixth International Conference on Information Science, Electrical, and Automation Engineering (ISEAE 2024) (Vol. 13275, pp. 250-255). SPIE.

[29] Zhu, J., Ortiz, J., & Sun, Y. (2024, November). Decoupled Deep Reinforcement Learning with Sensor Fusion and Imitation Learning for Autonomous Driving Optimization. In 2024 6th International Conference on Artificial Intelligence and Computer Applications (ICAICA) (pp. 306-310). IEEE.

[30] Lin, Y., Yao, Y., Zhu, J., & He, C. Application of Generative AI in Predictive Analysis of Urban Energy Distribution and Traffic Congestion in Smart Cities.

[31] Liu, Z., Costa, C., & Wu, Y. Expert Perception and Machine Learning Dimensional Risk Analysis.

[32] Sun, Y., Pargoo, N. S., Jin, P. J., & Ortiz, J. (2024). Optimizing Autonomous Driving for Safety: A Human-Centric Approach with LLM-Enhanced RLHF. arXiv preprint arXiv:2406.04481.

[33] Yang, J., Zhang, Y., Xu, K., Liu, W., & Chan, S. E. (2024). Adaptive Modeling and Risk Strategies for Cross-Border Real Estate Investments.

[34] Luo, D., Gu, J., Qin, F., Wang, G., & Yao, L. (2020, October). E-seed: Shape-changing interfaces that self drill. In Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology (pp. 45-57).