**Artificial Intelligence and Computing Innovations**

Intelligence Science Publishing

https://www.ispublishing.com

ISSN: 3070-4944

Vo.4, No.4, 2024

# *Heterogeneity-Aware Federated Optimization with Secure Aggregation for Distributed Data Mining*

**Yunjie Ke[1]***

[1]*Independent Researcher, San Francisco, USA*

**Corresponding author: yunjie.research@gmail.com*

***Abstract:*** This paper addresses the challenges of collaborative modeling in distributed data mining scenarios, stemming from difficulties in data centralization, strict privacy constraints, and significant client heterogeneity. It proposes an intelligent modeling framework integrating federated learning. This method uses communication rounds as its core organizational structure, enabling parameter-only collaborative training between a central server and multiple edge nodes. Clients define empirical risks based on local data and perform gradient updates, using regularization terms and proximal constraints to suppress local drift and improve optimization stability. The server performs weighted aggregation based on client sample proportions and the participating set, combining update increment metrics and pruning mechanisms to limit the impact of anomalous updates on the global model. Simultaneously, secure aggregation and update perturbation strategies are introduced to enhance the privacy protection and controllable leakage risk of uploaded information without exposing the original data. The method has a clear, modular, and composable overall structure, adaptable to varying client participation sizes and communication constraints, while balancing prediction performance and probabilistic reliability. Experiments are conducted under a unified evaluation protocol, comparing the proposed framework with several representative methods to verify its effectiveness and applicability in distributed data mining tasks.

***Keywords:*** Distributed data mining; privacy-preserving learning; secure aggregation; probability calibration

## 1. Introduction

As data elements increasingly penetrate fields such as financial risk control, healthcare, industrial internet, and smart cities, data mining models are becoming increasingly reliant on large-scale, multi-source data. However, in real-world applications, data is often scattered across different institutions and terminal devices, constrained by privacy compliance, business boundaries, and transmission costs, making it difficult to centrally aggregate it into a single computing center. Meanwhile, data generated at the edge possesses strong timeliness and high sensitivity; centralized collection and unified storage can

easily introduce additional security risks and governance costs[1]. Therefore, how to achieve cross-node collaborative modeling without aggregating raw data has become a key issue for the large-scale implementation of distributed data mining[2].

Federated learning provides a feasible path for collaborative training of distributed data, but it still faces several challenges when applied to real-world distributed data mining tasks. First, client-side data distribution often exhibits significant heterogeneity, with large differences in data size, category structure, and noise levels across different nodes, easily leading to local update biases and global optimization instability. Second, communication and computing power constraints mean that participating nodes may be intermittently online, with delays and inconsistencies in update uploads, further amplifying the volatility of the training process. Third, even without uploading raw data, model parameters or gradient updates may still carry exploitable information, potentially leading to privacy breaches during long-term iterations. These issues collectively constrain the reliability and controllability of federated learning in highly sensitive and heavily constrained data mining applications[3].

To address these pain points, distributed data mining intelligent modeling incorporating federated learning requires a more balanced technical solution that balances performance, stability, and privacy. On one hand, model training should be adaptable to heterogeneous clients, using reasonable local target design and global aggregation strategies to ensure more consistent learning directions and smoother convergence across different nodes[4]. On the other hand, the system needs updated anomaly suppression and controllable perturbation mechanisms to reduce the impact of extreme updates on the global model without sacrificing collaborative efficiency, while simultaneously enhancing the security of uploaded information. Furthermore, distributed data mining tasks are often used for risk decision-making and alarm classification; the probabilistic reliability of model output directly affects threshold selection and business processing. Therefore, the interpretability and consistency of prediction confidence should also be a core objective of the method design[5].

The main contributions of this paper are as follows: First, it proposes a federated collaborative modeling framework for distributed data mining, which organizes the client-side local optimization and server-side aggregation processes by communication rounds, enabling cross-node learning with only parameter interaction[6]. Second, it introduces a target modeling approach with regularization and proximal constraints on the client side to alleviate local drift caused by heterogeneous data and improve training stability. Third, it constructs a weighted aggregation and update pruning mechanism on the server side to suppress abnormal update amplitudes and enhance the robustness of global updates[7]. Fourth, it integrates secure aggregation and update perturbation strategies to reduce the risk of potential information leakage and promote more reliable probability output, thereby providing practical technical support for privacy-friendly intelligent mining in highly sensitive scenarios.

Recent studies on financial causal discovery, risk detection, robust data mining, and graph-based decision optimization further broaden the methodological context for privacy-preserving distributed data mining. Explainable cross-market causal discovery via financial knowledge graphs demonstrates that domain graphs can organize heterogeneous market entities and causal signals, supporting interpretable modeling in high-risk cross-institutional scenarios[8]. Microstructure-aware risk detection in high-frequency trading using embedding and sequence modeling highlights the need to capture fine-grained temporal risk signals from sensitive financial data without exposing raw observations[9]. A unified semi-supervised and robust optimization framework for low-resource noisy data mining provides a methodological basis for learning under limited labels, heterogeneous noise, and unstable client distributions[10]. A differentiable mixed-integer optimization framework for social graph modeling and node decision-making extends graph-based decision logic to structured optimization, offering useful insight for client selection, update pruning, and risk-aware aggregation in federated environments[11]. Together, these studies motivate a framework that combines interpretable structural knowledge, sequence-level risk representation, robust optimization, and privacy-preserving collaborative learning.

## 2. Method

### 2.1 Overall Framework

This research addresses the distributed data mining and intelligent modeling task that integrates federated learning. It constructs an end-to-end privacy-friendly collaborative training framework: multiple rounds of communication iteration occur between a central server and multiple edge clients.

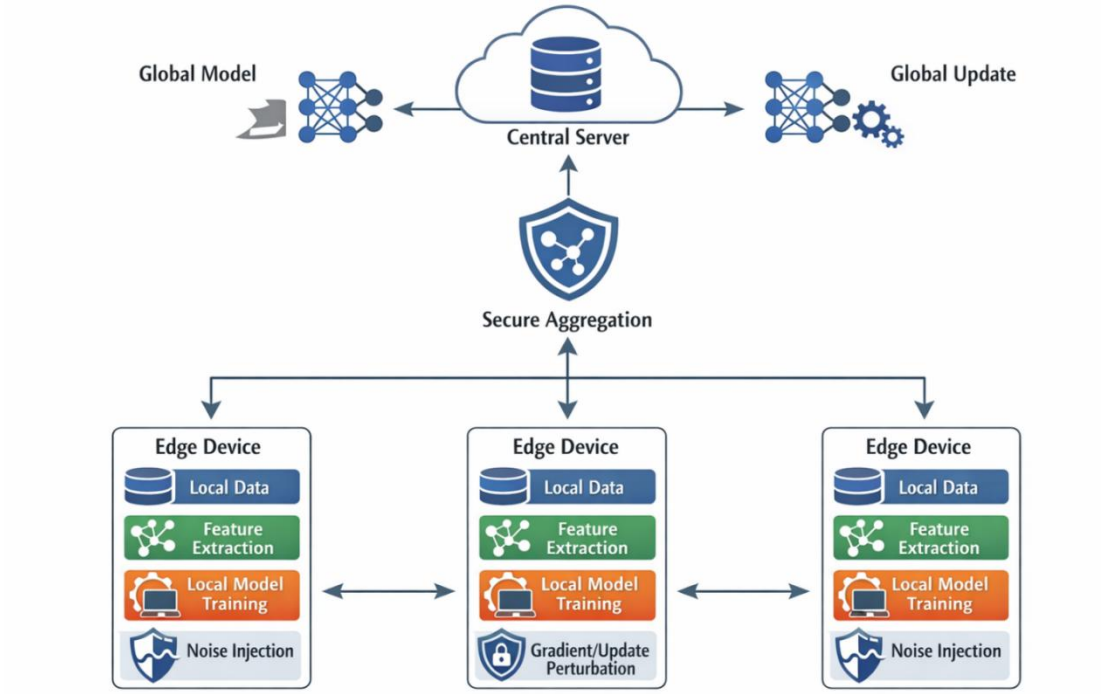



Figure 1. This diagram illustrates the overall architecture of distributed data mining intelligent modeling that integrates federated learning, including global model maintenance and updates on a central server, and local data processing and model training processes on multiple edge devices. The system achieves cross-client parameter fusion without sharing original data through secure aggregation and update perturbation mechanisms, thus enabling privacy-friendly and scalable collaborative modeling.

Each client completes feature extraction and model updates only on its local data, while the server only receives and aggregates model parameters or gradients, thereby continuously optimizing the global model without collecting raw data. Let the global participant set be:

$$\mathcal{C} = \{1,2,\ldots,K\} \tag{1}$$

where $K$ denotes the number of clients, and $k \in \mathcal{C}$ identifies the $k$ th client. The $k$ th client holds a local dataset:

$$\mathcal{D}_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k} \tag{2}$$

In Formula (2), $n_k$ is the number of samples on client $k$, $x_{k,i}$ denotes the input features of the $i$ th sample, and $y_{k,i}$ is the supervision signal or target label for the mining task. The global data size is defined as:

$$N = \sum_{k=1}^{K} n_k \tag{3}$$

This quantity characterizes the sum of sample counts across all clients. To quantify the contribution of each client to aggregation, the sample proportion weight is introduced as:

$$p_k = \frac{n_k}{N} \tag{4}$$

and we have $p_k \in [0,1]$ and $\sum_{k=1}^{K} p_k = 1$. Let the global model parameters be:

$$w \in \mathbb{R}^d \tag{5}$$

where $d$ is the parameter dimension, and the global parameters at communication round $t$ are denoted by $w^t$.

At the beginning of each round, the server broadcasts $w^t$ to participating clients. Each client performs local updates to obtain $w_k^{t+1}$ and uploads it, and the server aggregates them to produce the next round global parameters. With communication rounds as the time axis, this mechanism enables the distributed data mining model to stably converge to a usable global representation under heterogeneous data distributions, heterogeneous computing resources, and privacy constraints.

## 2.2 Local Data Mining Objective and Optimization

On client $k$, a local objective function $F_k(w)$ is defined to uniformly describe the empirical risk for data mining tasks such as classification, regression, or clustering consistency. First, the single sample loss is given by:

$$l = f(x; w, y) \tag{6}$$

In Formula (6), $f(x; w)$ denotes the model prediction output for input $x$ under parameters $w$, $l$ is the loss function that matches the mining objective, and $y$ is the supervision signal or pseudo label. The empirical risk of client $k$ is defined as:

$$F_k(w) = \frac{1}{n_k} \sum_{i=1}^{n_k} l\left(f\left(x_{k,i}; w\right), y_{k,i}\right) \tag{7}$$

This summation iterates over all local samples on client $k$. To suppress overfitting and improve generalization, an $L_2$ regularization term is added to obtain the regularized objective.

$$F_k^{\text{reg}}(w) = F_k(w) + \frac{\lambda}{2} \| w\|_2^2 \tag{8}$$

In this notation, $\lambda \geq 0$ is the regularization strength and $\|\cdot\|_2$ is the Euclidean norm. The client adopts gradient based local optimization, and the gradient of the regularized objective is written as:

$$g_k^t = \nabla_w F_k^{\text{reg}}\left(w^t\right) \tag{9}$$

The operator $\nabla_w$ denotes the gradient with respect to parameters $w$, and $g_k^t$ denotes the gradient direction computed by client $k$ at round $t$. A single local gradient descent update can be written as:

$$w_k^{t+1} = w^t - \eta\, g_k^t \tag{10}$$

where $\eta > 0$ is the learning rate and $w_k^{t+1}$ is the locally updated parameters of client $k$.

Considering the non independent and identically distributed characteristics that commonly exist in distributed data mining, a proximal constraint to the global model can be introduced to obtain a more stable form of local optimization:

$$F_k(w) = F_k(w) + \frac{\mu}{2} \| w - w^t\|_2^2 \tag{11}$$

where $\mu \geq 0$ is the proximal coefficient, and $\| w - w^t\|_2^2$ constrains local updates from drifting too far away from the current global model, thereby reducing client drift and enhancing cross client consistency. The pseudocode flow of the local objective and optimization is shown in Figure 2.

**Algorithm 1** Client-$k$ Local Optimization with $\ell_2$ Regularization and Proximal Constraint

**Require:** Client dataset $\mathcal{D}_k = \{(x_{k,i}, y_{k,i})\}_{i=1}^{n_k}$, global model $w^t$, learning rate $\eta$, reg. strength $\lambda \geq 0$, proximal coeff. $\mu \geq 0$

**Ensure:** Updated local model $w_k^{t+1}$

1: Initialize $w \leftarrow w^t$

2: Define single-sample loss $\ell(f(x; w), y)$

3: Define empirical risk:

$$F_k(w) \leftarrow \frac{1}{n_k} \sum_{i=1}^{n_k} \ell\big(f(x_{k,i}; w), y_{k,i}\big)$$

4: Define local objective (regularized + proximal):

$$F_k^{\text{loc}}(w) \leftarrow F_k(w) + \frac{\lambda}{2}\|w\|_2^2 + \frac{\mu}{2}\|w - w^t\|_2^2$$

5: Compute gradient $g_k^t \leftarrow \nabla_w F_k^{\text{loc}}(w)$

6: Update $w_k^{t+1} \leftarrow w - \eta\, g_k^t$

7: **return** $w_k^{t+1}$

Figure 2. Pseudocode Flowchart of Local Data Mining Objective and Optimization

## 2.3 Federated Aggregation and Robust Update

The server employs a weighted averaging strategy to fuse model updates uploaded by clients, thereby forming a globally consistent mining model. The standard weighted aggregation can be expressed as:

$$w^{t+1} = \sum_{k=1}^{K} p_k \; w_k^{t+1} \tag{12}$$

$p_k$ is the sample proportion weight and $w^{t+1}$ denotes the global parameters at round $t+1$. To reduce communication overhead and improve scalability, only a subset of clients can be selected to participate in each round, denoted by $S^t \subseteq \mathcal{C}$. In this case, the aggregation can be rewritten as:

$$w^{t+1} = \sum_{k \in S^t} \alpha_k^t \; w_k^{t+1} \tag{13}$$

To handle noisy updates or anomalous clients that may occur in distributed data mining, an update magnitude measure is introduced as:

$$\Delta_k^t = w_k^{t+1} - w^t \tag{14}$$

This increment describes the change of client $k$ relative to the global parameters at round $t$. Based on the increment, a simple robust clipping can be performed as:

$$\Delta_k^t = \Delta_k^t \cdot \min\left(1, \frac{\tau}{\|\Delta_k^t\|_2}\right) \tag{15}$$

where $\tau > 0$ is the clipping threshold, and the update is scaled to limit the destabilizing effect of excessively large steps on the global model. Then, the global update can be written as:

$$w^{t+1} = w^t + \sum_{k \in S^t} \alpha_k^t \; \Delta_k^t \tag{16}$$

The first term on the right hand side is the global parameters from the previous round, and the second term is the aggregated clipped increment, which suppresses abnormal updates while retaining normal updates.

### 2.4 Privacy Preservation and Secure Communication

To further enhance privacy protection without exposing the original data, the gradient or parameter increments uploaded by the client can be perturbed. Taking incremental form as an example, Gaussian noise can be defined as:

$$z_k^{(t)} \sim \mathcal{N}(0, \sigma^2 I) \tag{17}$$

Where $\mathcal{N}(0, \sigma^2 I)$ represents a multidimensional Gaussian distribution with mean 0 and covariance $\sigma^2 I$, $\sigma > 0$ represents the noise intensity, $I$ is the identity matrix, and $z_k^{(t)}$ is the random noise vector generated by client $k$ in round $t$. The uploaded message is obtained by perturbing the pruned increment.

$$\widetilde{\Delta}_k^{(t)} = \widehat{\Delta}_k^{(t)} + z_k^{(t)} \tag{18}$$

In Formula (18), $\widetilde{\Delta}_k^{(t)}$ is the increment after adding noise, used to reduce the risk of individual samples being identifiable in updates. The perturbation increment is used during server-side aggregation to obtain:

$$w^{(t+1)} = w^{(t)} + \sum_{k \in S^{(t)}} \alpha_k^{(t)} \widetilde{\Delta}_k^{(t)} \tag{19}$$

Here, $\widetilde{\Delta}_k^{(t)}$ statistically preserves the overall update direction while enhancing privacy through a noise mechanism. To facilitate a unified description of the end-to-end training objective, the global objective function is defined as follows:

$$F(w) = \sum_{k=1}^{K} p_k F_k(w) \tag{20}$$

$F_k(w)$ represents the client's empirical risk, $p_k$ represents the weight, and $F(w)$ represents the globally weighted risk. Based on the above definitions, the overall training process can be abstracted as iterative optimization of $F(w)$ under privacy protection and communication constraints: the client minimizes $F_k(w)$ on local data, and the server constructs the evolutionary sequence of wt through weighted aggregation, realizing collaborative learning and controllable privacy exposure in distributed data mining intelligent modeling.

## 3. Experimental Results and Analysis

### 3.1 Dataset

This study selects FEMNIST from the LEAF benchmark as an open-source dataset to characterize distributed data mining modeling scenarios incorporating federated learning. FEMNIST is a federated version of the EMNIST handwritten character data. Its core feature is that it divides samples into multiple clients based on the writer's identity, naturally forming a cross-client, non-independent, identically distributed data structure that can reflect issues such as data skew and sample imbalance in real-world distributed environments. The LEAF framework provides complete data generation, preprocessing, and federated partitioning tools for this dataset, facilitating the reproduction of the "central server + multiple edge devices" training and aggregation process.

In FEMNIST, each client corresponds to a writer, and its local data consists of handwritten letter and number samples from that writer. The task is typically set as multi-class character recognition, with grayscale handwritten images as input and character category labels as output. The federated partitioning approach results in significant differences in character distribution, writing style, and sample size among different clients. It is suitable for verifying the adaptability and scalability of key modules in distributed data mining modeling, such as local target optimization, client sampling participation, server weighted aggregation, update pruning, noise perturbation, and safe aggregation.

### 3.2 Dataset preprocessing

To ensure controllable data distribution and reproducible training process across clients in a federated learning scenario, the raw data is first uniformly cleaned and formatted: damaged samples and missing label records are removed to ensure that each sample contains valid input and corresponding category labeling; then, the input images are uniformly converted to single-channel grayscale and their size is normalized to ensure that all samples meet the same spatial resolution requirements; linear normalization is performed at the pixel intensity level to scale the values to a fixed range for stable optimization; while keeping the client identity classification unchanged, the samples within each client are randomly shuffled and divided into local training and local validation sets according to a fixed ratio for local updates and early stopping monitoring on the client side; finally, the sample size of each client is counted and the aggregation weight is calculated to ensure that the subsequent server-side weighted aggregation and client sampling strategies are consistent with the data scale.

This work focuses on the common challenges of distributed data mining, namely, addressing the modeling degradation caused by heterogeneous and non-IID client distributions, limited communication, and privacy protection under the constraint of data not leaving the domain. To avoid overly binding experimental conclusions to a specific industry, the financial risk control, healthcare, industrial internet, and smart cities mentioned in the introduction are only used as potential application backgrounds to illustrate the real-world prevalence of these common challenges, rather than direct empirical commitments to specific business data formats. In the experimental design, we chose FEMNIST as a classic benchmark for federated learning and distributed data mining. Its significant client-level distribution bias can rigorously test the effectiveness and stability of the proposed federated modeling mechanism in non-IID and heterogeneous environments under controllable conditions. It is important to emphasize that although this paper uses classification tasks as the main validation vehicle, the proposed modeling mechanism and training strategy are essentially geared towards decentralized optimization and robust representation learning, and can be seamlessly transferred to a wider range of data mining tasks, such as tabular, time series, and multimodal data mining.

### 3.3 Experimental Results and Analysis

To verify the effectiveness of the proposed integrated federated learning distributed data mining modeling framework under conditions of privacy protection, communication constraints, and data heterogeneity, this paper selects several representative federated optimization and secure aggregation methods as comparative baselines and aligns the metrics under a unified data partitioning and evaluation protocol. The comparative settings cover key research directions such as classical federated averaging, heterogeneous optimization robustness, client drift correction, adaptive federated optimization, normalized aggregation, and secure aggregation, in order to comprehensively characterize the methods from dimensions such as prediction performance and probability calibration. The experimental results are shown in Table 1.

Overall, the baseline methods showed a steady increase in performance across six metrics: Accuracy, Precision, Recall, F1, AUROC, and AUPRC. This indicates that improvements in federated optimization strategies and aggregation mechanisms have continuously enhanced the models' discriminative capabilities on distributed heterogeneous data. Ours achieved the highest levels across all six metrics, with Accuracy at 90.65, F1 at 90.21, AUROC at 94.71, and AUPRC at 93.84. These represent improvements of 2.25, 2.44, 2.55, and 2.42 percentage points, respectively, compared to the strongest baseline, Zhang et al. This demonstrates that the proposed method exhibits consistent advantages in overall correctness, positive class identification quality, and threshold-independent ranking discriminative ability, and that the improvements are not concentrated on a single metric.

Regarding probabilistic reliability, ECE and Brier continuously decreased with method enhancement, reflecting a gradual improvement in calibration quality. Ours achieved an ECE of 2.71, while Brier

achieved 0.078, representing further reductions of 1.11 and 0.023 compared to Zhang et al., indicating that while maintaining stronger classification performance, the predicted probabilities are more consistent with the true accuracy, and the probability error is smaller. Overall, these results cover both "higher performance" and "more reliable" benefits, meaning they not only improve accurate discrimination but also reduce the risks associated with overconfidence or unstable probability outputs, better meeting the robustness and interpretability requirements of distributed data mining tasks.

Table 1. Experimental results compared with other models

| Method | Accuracy | Precision | Recall | F1 | AUROC | AUPRC | ECE(%) | Brier |
|---|---|---|---|---|---|---|---|---|
| McMahan et al.[12] | 82.34 | 81.92 | 81.41 | 81.66 | 87.15 | 86.33 | 5.62 | 0.132 |
| Li et al.[13] | 84.10 | 83.75 | 83.21 | 83.48 | 88.46 | 87.72 | 5.21 | 0.126 |
| Karimireddy et al.[14] | 85.56 | 85.10 | 84.63 | 84.86 | 89.70 | 88.94 | 4.88 | 0.119 |
| Reddi et al.[15] | 86.27 | 85.83 | 85.34 | 85.58 | 90.24 | 89.39 | 4.61 | 0.114 |
| Wang et al.[16] | 87.14 | 86.72 | 86.20 | 86.46 | 91.02 | 90.27 | 4.29 | 0.108 |
| Bonawitz et al.[17] | 87.73 | 87.39 | 86.90 | 87.14 | 91.55 | 90.74 | 4.05 | 0.104 |
| Zhang et al.[18] | 88.40 | 88.03 | 87.52 | 87.77 | 92.16 | 91.42 | 3.82 | 0.101 |
| Ours | 90.65 | 90.41 | 90.02 | 90.21 | 94.71 | 93.84 | 2.71 | 0.078 |

In the communication iteration process of federated learning, the gradient updates generated by local optimization exhibit significant statistical distribution changes with each round. Visualizing the gradient norm and parameter update magnitude at the distribution level along the communication round dimension helps to intuitively characterize the impact of training stability, client drift, and aggregation constraints on the update behavior. To highlight the updated characteristics of our proposed method, a ridge plot is used to hierarchically display the distribution pattern across different rounds.

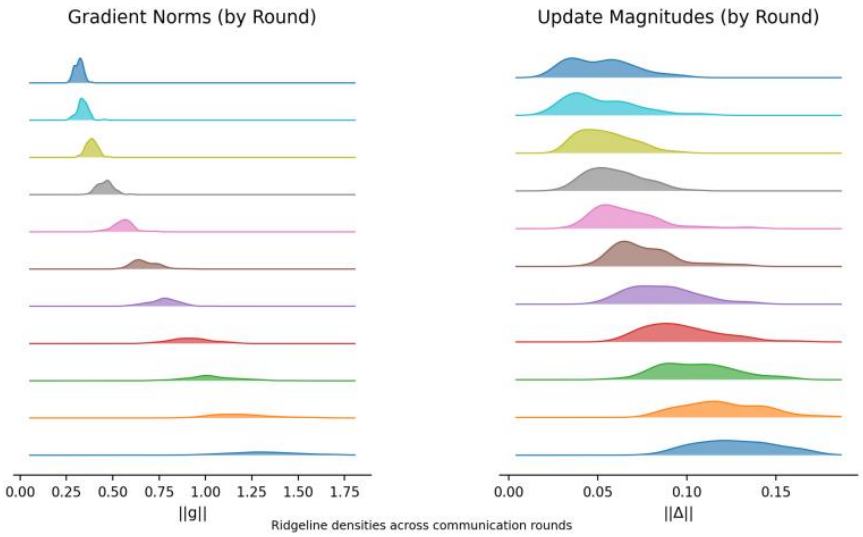



Figure 3. A ridge plot visualization experiment of the gradient norm and update magnitude distribution of communication rounds.

As shown in Figure 3, with each communication round, the gradient norm distribution gradually converges to a more concentrated and stable form, with little change in the main peak position and significant compression of the long tail. This indicates that during the federated iteration process, the

proposed method can continuously suppress the drastic fluctuations caused by client heterogeneity and local optimization, making the training process smoother, the update direction more consistent, and exhibiting good optimization stability and convergence characteristics overall.

The updated amplitude distribution also shows a smooth and controllable evolution trend. The density pattern of each round is relatively regular, without abnormal surges or distribution fragmentation, and tail expansion is effectively limited. This phenomenon shows that, under the combined effect of aggregation and perturbation mechanisms, the proposed method can maintain sufficient effective updates to drive global model improvement while avoiding instability caused by excessive parameter jumps, thus obtaining more reliable training dynamics and better overall performance.

The strength of the safety perturbation directly affects the randomness level of the uploaded updates during federated training, thereby altering the global optimization trajectory and the shape of the prediction probability. By treating the perturbation strength as a continuous control variable and observing the linkage between model performance and probabilistic reliability using a phase diagram, the usable range and inflection point positions can be visually presented. This visualization focuses solely on the method presented in this paper, used to verify whether the training dynamics under different perturbation configurations are sufficiently smooth and interpretable; the experimental results are shown in Figure 4.

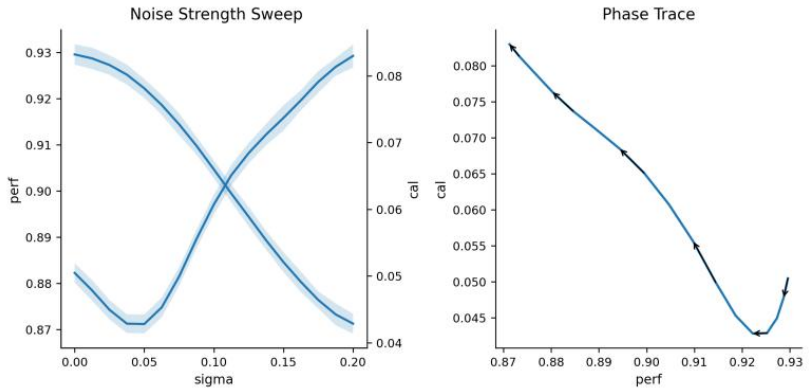



Figure 4. Noise strength sweep and performance-calibration trajectory. The left panel reports performance (Perf, left y-axis) and calibration error (Cal, right y-axis) versus the noise scale $\sigma$. The right panel shows the corresponding Perf-Cal trajectory, where arrows indicate increasing $\sigma$

Furthermore, as can be seen from the biaxial curves on the left, the model performance and calibration error exhibit a clear and smooth correlation as the strength of the safety perturbation changes. The overall trend is continuous, without any abrupt changes or unstable oscillations. This indicates that the proposed method is highly adaptable to the randomness introduced by the perturbation, maintains a controllable optimization trajectory during training, and demonstrates stable output characteristics under different perturbation configurations.

The phase diagram on the right further condenses this correlation into a coherent trajectory. The curve direction is clear, and the shape is regular, indicating that there is an interpretable path of change between performance and calibration, and that a good balance can be maintained simultaneously within a certain range. Overall, the proposed method demonstrates good robustness and adjustability under safety perturbations, facilitating the selection of appropriate perturbation intensity and showcasing the advantages of more reliable probabilistic output and more credible training dynamics.

## 4. Conclusion

This paper addresses the problem of intelligent modeling in distributed data mining using federated learning. For real-world scenarios characterized by decentralized data, strong privacy constraints, and significant client heterogeneity, it proposes an end-to-end collaborative learning framework. This framework uses the communication iteration between a central server and multiple edge nodes as the main thread. In the local optimization phase, regularization and proximal constraints are introduced to mitigate client drift. In the server aggregation phase, weighted fusion combined with update pruning enhances robustness. Simultaneously, secure aggregation and perturbation mechanisms reduce the risk of update leakage, thus achieving scalable distributed modeling without sharing original data. The overall method has a clear structure, composable modules, and controllable engineering implementation costs, making it suitable for various data mining tasks and diverse device environments.

From an application value perspective, this work provides a more usable technical path for data-driven modeling in cross-institutional collaborations, unlocking the joint value of dispersed data while maintaining compliance. For fields such as financial risk control, anti-fraud, and cross-platform user behavior mining, the method can support multiple parties in building more robust discriminative models without exposing sensitive transaction and identity information, and provides more reliable probabilistic expressions in terms of model output reliability. For highly sensitive data scenarios such as healthcare and public health monitoring, this method promotes multi-center collaborative modeling, reducing the limitations of data silos on model generalization. For edge-intensive systems such as the Industrial Internet and smart cities, the method maintains training stability under conditions of weak connectivity, uneven computing power, and data distribution drift, providing continuous updates for online decision-making and anomaly detection.

The significance of this method lies not only in improving the feasibility of distributed data mining but also in advancing collaborative modeling from usability to reliability. By constraining and protecting the local update magnitude and global aggregation process, model training dynamics become smoother and update behavior more interpretable, helping to reduce the instability risks and uncontrollable biases common in federated training. Simultaneously, calibration mechanisms make the model's probability output closer to the true expression of uncertainty, which is particularly crucial for high-risk decision-making scenarios, providing a more reliable basis for subsequent threshold selection, alarm classification, and human-machine collaborative review. Overall, this work provides a more balanced engineering solution among privacy, stability, and scalability, offering direct reference value for the practical deployment of distributed intelligent modeling.

Future work can be further advanced around more complex distributed constraints and broader task forms. On the one hand, it can explore more efficient communication organization and resource scheduling strategies for larger-scale clients and more heterogeneous network environments, enabling the model to iterate efficiently even under low bandwidth and intermittent connection conditions. On the other hand, the framework can be extended to multimodal data and time-series streaming data mining, supporting cross-device and cross-platform joint representation learning and continuous learning needs. Simultaneously, without compromising privacy principles, further strengthening the ability to identify and suppress abnormal clients and noisy updates will help improve the reliability of large-scale real-world deployments. With the improvement of privacy-preserving computing and edge intelligent infrastructure, the approach proposed in this paper is expected to play a role in more industry collaboration scenarios, promoting data sharing from the data layer to the model layer, and driving the large-scale deployment of intelligent mining capabilities along a compliant, secure, and sustainable path.

## References

[1] Gao L, Fu H, Li L, et al. Feddc: Federated learning with non-iid data via local drift decoupling and correction[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 10112-10121.

[2] Guo Y, Tang X, Lin T. Fedbr: Improving federated learning on heterogeneous data via local learning bias reduction[C]//International conference on machine learning. PMLR, 2023: 12034-12054.

[3] Jiang X, Rodomanov A, Stich S U. Federated optimization with doubly regularized drift correction[J]. arXiv preprint arXiv:2404.08447, 2024.

[4] Guo H, Wang H, Song T, et al. Siren+: Robust Federated Learning With Proactive Alarming and Differential Privacy[J]. IEEE Transactions on Dependable and Secure Computing, 2024, 21(5): 4843-4860.

[5] Campos E M, Gonzalez-Vidal A, Hernández-Ramos J L, et al. FedRDF: A robust and dynamic aggregation function against poisoning attacks in federated learning[J]. IEEE Transactions on Emerging Topics in Computing, 2024, 13(1): 48-67.

[6] Hasan M, Zhang G, Guo K, et al. Calibrated one round federated learning with bayesian inference in the predictive space[C]//Proceedings of the AAAI conference on artificial intelligence. 2024, 38(11): 12313-12321.

[7] Wang Y, Zhang A, Wu S, et al. VOSA: Verifiable and oblivious secure aggregation for privacy-preserving federated learning[J]. IEEE Transactions on Dependable and Secure Computing, 2022, 20(5): 3601-3616.

[8] Chen L, Liu S. Explainable Cross-Market Causal Discovery via Financial Knowledge Graphs[J]. 2024, 3(1).

[9] Su Z. Microstructure-Aware Risk Detection in High-Frequency Trading Using Embedding and Sequence Modeling[J]. 2024, 4(3).

[10] Mei Y. Unified Semi-Supervised and Robust Optimization Framework for Low-Resource Noisy Data Mining[J]. Journal of Computer Technology and Software, 2024, 3(7).

[11] Ren X. A Unified Framework for Social Graph Modeling and Node Decision-Making via Differentiable Mixed-Integer Optimization[J].2024, 3(5).

[12] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Artificial intelligence and statistics. Pmlr, 2017: 1273-1282.

[13] Li T, Sahu A K, Zaheer M, et al. Federated optimization in heterogeneous networks[J]. Proceedings of Machine learning and systems, 2020, 2: 429-450.

[14] Karimireddy S P, Kale S, Mohri M, et al. Scaffold: Stochastic controlled averaging for federated learning[C]//International conference on machine learning. PMLR, 2020: 5132-5143.

[15] Reddi S, Charles Z, Zaheer M, et al. Adaptive federated optimization[J]. arXiv preprint arXiv:2003.00295, 2020.

[16] Wang J, Liu Q, Liang H, et al. Tackling the objective inconsistency problem in heterogeneous federated optimization[J]. Advances in neural information processing systems, 2020, 33: 7611-7623.

[17] Bonawitz K, Ivanov V, Kreuter B, et al. Practical secure aggregation for privacy-preserving machine learning[C]//proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security. 2017: 1175-1191.

[18] Kairouz P, McMahan H B, Avent B, et al. Advances and Open Problems in Federated Learning[J]. Foundations and Trends® in Machine Learning, 2021, 14(1-2): 1-210.