Journal of Computer Technology and Software

ISSN: 2998-2383

Vol. 3, No. 4, 2024

Differential Privacy-Enhanced Federated Learning for Robust AI Systems

Yilin Li
Carnegie Mellon University, Pittsburgh, USA
ireneli961111@gmail.com

Abstract: This paper proposes a differential-privacy-enhanced federated learning framework to address the challenges of privacy protection and robustness in federated learning. The study first analyzes the limitations of traditional federated learning under parameter aggregation and distribution heterogeneity, noting that relying solely on distributed modeling is insufficient to prevent data leakage and adversarial risks. In the method design, gradient clipping and noise injection are introduced to enforce differential privacy, and robust aggregation operators are employed to suppress negative impacts from malicious clients or abnormal distributions. On this basis, the framework is systematically evaluated through comparative and sensitivity experiments across dimensions such as learning rate, client sampling rate, data imbalance, and adversarial noise amplitude, using accuracy, precision, recall, and F1-Score as evaluation metrics. The results show that the proposed method maintains high utility while ensuring privacy and demonstrates stable performance in complex environments. This work not only validates the effective integration of differential privacy and robustness design but also provides a complete technical pathway for building trustworthy intelligent systems in high-risk and sensitive data scenarios. Based on this background, the integration of differential privacy and federated learning has become a research focus in recent years, as studies show that introducing differential privacy into distributed modeling can protect user data while improving system reliability under non-ideal conditions. Such integration can resist external attacks and suppress interference from malicious clients, thereby enhancing overall robustness. However, most existing work still emphasizes either privacy protection or robustness in isolation, lacking a systematic framework to optimize both simultaneously. Therefore, exploring differential-privacy-enhanced federated learning to construct more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

Keywords: Differential privacy; federated learning; robustness; sensitivity experiments

1. Introduction

In the era of big data and artificial intelligence, data-driven intelligent systems have shown great potential in healthcare, finance, transportation, government, and industry. However, the wide collection and use of data have brought serious concerns about user privacy and data security. Traditional centralized modeling relies on aggregating large amounts of raw data on a central server[1]. This approach increases the risk of data leakage and creates compliance and ethical challenges. At the same time, AI systems often appear fragile in complex environments with distribution shifts, noise, and malicious attacks. Ensuring both privacy protection and robust performance has become a critical issue[2].

Federated learning, as an emerging distributed training paradigm, provides a new solution. By moving model training to local devices, federated learning enables collaborative modeling across devices or institutions without uploading raw data. This mechanism reduces the risk of data leakage and supports data compliance and cross-domain cooperation. Yet federated learning also faces challenges. Data heterogeneity among local devices can lead to unstable training. In addition, parameter aggregation and transmission are still vulnerable to attacks and leaks. Relying on traditional federated learning alone cannot achieve an ideal balance between privacy and robustness[3].

Differential privacy, as one of the most influential privacy protection techniques, brings stronger security guarantees to federated learning. By adding carefully designed noise to gradients or model parameters, differential privacy limits the impact of any single sample on the final model. This ensures that attackers cannot infer sensitive information through reverse analysis[4]. The mechanism strengthens privacy control within federated learning. However, the introduction of noise also reduces utility. In high-dimensional data or complex model settings, strong noise can weaken performance. Balancing privacy constraints with model utility and robustness has therefore become a shared focus in both academia and industry.

From the perspective of robustness, AI systems must withstand not only conventional noise and distribution imbalance, but also adversarial examples, parameter tampering, and malicious client uploads. Existing approaches often trade accuracy for security or strengthen robustness while neglecting privacy. The integration of differential privacy and federated learning offers a complementary solution. Differential privacy reduces the risk of individual client attacks and improves

system security. At the same time, the distributed structure of federated learning provides flexible strategies for noise allocation and control, allowing the model to remain stable and adaptive in complex environments[5].

In summary, differential-privacy-enhanced federated learning for robust AI systems has significant theoretical and practical value. On one hand, it promotes the integration of privacy protection and intelligent modeling, addressing urgent demands for data security and compliance[6,7]. On the other hand, it provides reliable technical support for AI systems in high-risk fields such as financial risk control, smart healthcare, and smart cities. In future intelligent development, this direction is expected to become a key path for safeguarding user rights and enhancing AI trustworthiness, thus laying a foundation for the healthy and sustainable growth of artificial intelligence.

2. Related work

In the research field that intersects privacy protection and artificial intelligence, federated learning is regarded as an important framework that balances data use and privacy[8]. Compared with traditional centralized training, it avoids centralized storage of raw data through distributed parameter sharing. This reduces the risks of data leakage and compliance violations. Many studies focus on the optimization of federated learning, including improving communication efficiency, ensuring stability under heterogeneous data, and achieving fairness in cross-domain collaboration. These studies have laid the foundation for large-scale distributed intelligent systems. However, in practical applications, relying solely on federated learning is insufficient to defend against complex privacy threats and adversarial risks[9].

To compensate for the limitations of federated learning in privacy protection, differential privacy has been introduced into related research. By adding noise during parameter updates or gradient uploads, differential privacy minimizes the influence of any single data sample. It provides strong privacy guarantees from a theoretical perspective[10]. Researchers have explored various mechanisms, including noise granularity control, adaptive privacy budget allocation, and utility preservation in high-dimensional settings. These improvements have greatly enhanced the applicability of differential privacy in real tasks. Yet the integration of differential privacy with federated learning also brings performance degradation. Balancing privacy protection and model utility has therefore become the central challenge in current research[11].

At the same time, research on the robustness of AI systems has been increasing. Traditional deep learning models are often vulnerable when facing out-of-distribution data, adversarial attacks, and noise[12]. To address this, the academic community has proposed many strategies, such as regularization methods, adversarial training, and multimodal data fusion. These methods improve model stability and generalization to some extent. However, they usually fail to address privacy constraints. In high-risk or sensitive data scenarios, focusing only on robustness while neglecting privacy can leave systems exposed to security risks.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies show that introducing differential privacy into distributed modeling can protect user data while improving system reliability in non-ideal environments. This integration can defend against external attacks and suppress interference from malicious clients, thereby enhancing robustness to some extent[13]. However, most existing work still takes a single perspective, often emphasizing either privacy or robustness. There is a lack of systematic frameworks that optimize both dimensions simultaneously. Exploring differential-privacy-enhanced federated learning to build more robust AI systems is thus not only an extension of existing research but also a necessary direction for promoting trustworthy artificial intelligence.

3. Proposed Approach

When building a robust differentially private federated learning framework, we first need to clarify the basic process of federated learning. Assume there is a global model parameter vector θ. In round t of training, the central server distributes it to the set of participating clients C_t. Each client i uses its local data distribution D_i to optimize the model parameters. The objective function can be formalized as weighted empirical risk minimization:

$$\min F(\theta) = \sum_{i=1}^{N} \frac{|D_i|}{\sum_{j=1}^{N} |D_j|} f_i(\theta) \tag{1}$$

Where $f_i(\theta)$ represents the local loss function of client i. To ensure the effectiveness of distributed collaboration, each client updates the parameters locally by gradient descent:

$$\theta_i^{t+1} = \theta_t - \eta \nabla f_i(\theta^t) \tag{2}$$

After completing several iterations, the updated results are uploaded to the central server for aggregation. The overall model architecture is shown in Figure 1.

In the process of ensuring privacy, the introduction of a differential privacy mechanism is the core link. Specifically, when uploading parameters in each round, the gradient needs to be clipped to control its sensitivity:

$$g_i^t = \frac{g_i^t}{\max(1, \frac{\| g_i^t \|_2}{C})} \tag{3}$$

Where g_i^t represents the client's local gradient and C is the clipping threshold. On this basis, differential privacy is achieved by adding Gaussian noise to the clipped gradient:

$$g_i^t = \hat{g}_i^t + N(0, \sigma^2 C^2 I) \tag{4}$$

Here, σ controls the noise intensity, effectively balancing privacy budget and model performance. During the aggregation phase, the server updates the perturbed gradients to ensure the privacy and security of global parameters.

To further enhance the robustness of the system, considering that the client may have malicious uploads or abnormal distribution, the aggregation method uses a robust aggregation operator instead of a simple weighted average. The robust aggregation function is defined as:

$$\theta^{t+1} = A(\{g_i^t \mid i \in C_t\}) \tag{5}$$

Where $A(\cdot)$ represents a robust aggregation mechanism, such as median-based or truncated mean-based methods, to mitigate the impact of anomalous updates. Within this framework, the global update process effectively improves resilience to noise, adversarial attacks, and heterogeneous distributions while maintaining differential privacy constraints.

Finally, under the theoretical constraints of differential privacy, the overall privacy budget of the system can be defined by the privacy loss accumulation formula. Assuming that the differential privacy parameter of each round is (ε, δ), after T rounds of iterations, the overall privacy budget can be obtained using the privacy accounting mechanism:

$$\varepsilon_{total} = \sqrt{2T\ln(\frac{1}{\delta})} \cdot \varepsilon + T \cdot \varepsilon(e^{\varepsilon} - 1) \tag{6}$$

This expression provides theoretical privacy guarantees for the algorithm over multiple iterations, thus providing a solid security boundary for practical deployment. By organically combining differential privacy with a robust aggregation strategy, this method can achieve both privacy protection and robustness enhancements in large-scale distributed environments, laying a methodological foundation for building trustworthy AI systems.

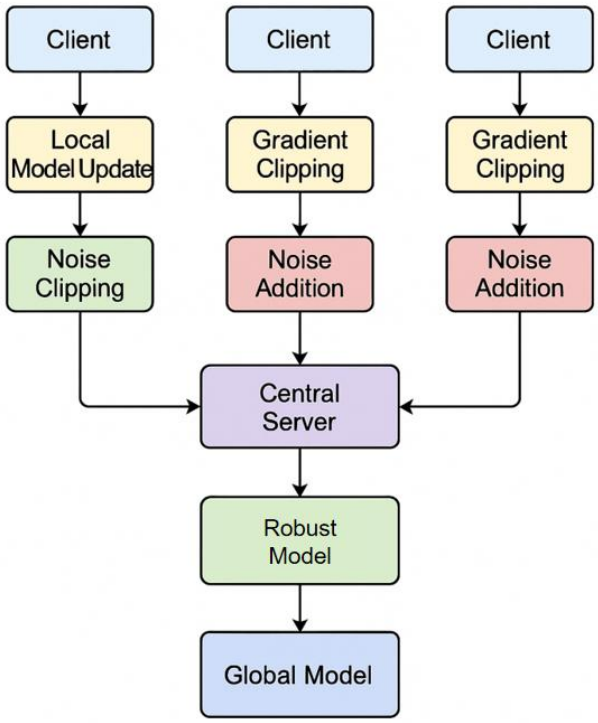

Figure 1. Overall model architecture

4. Performance Evaluation

4.1 Dataset

In this study, the dataset used is the FEMNIST dataset from the LEAF benchmark. This dataset is an extended version of the handwritten character recognition task. It is built on classical handwritten digits and letters and is repartitioned for federated learning scenarios. The data are assigned to different clients, with each client corresponding to the set of characters written by one user. This naturally leads to non-independent and non-identical distributions, which align with the heterogeneity assumption in federated learning.

The dataset contains 62 classes, covering digits and both uppercase and lowercase letters. The total sample size exceeds 800,000 images. The data are stored as grayscale images with a size of 28 × 28 pixels. The dataset is lightweight and standardized, making it suitable for modeling and validation on resource-constrained devices and in large-scale distributed settings. Due to significant differences in data volume across clients and highly imbalanced class distributions, this dataset has become an important benchmark for evaluating federated learning algorithms under complex conditions.

The choice of this dataset is motivated by its ability to reflect real-world situations of uneven data distribution and sample imbalance. It also presents relatively high task difficulty and strong generality. The dataset has been widely used in federated learning research. It provides a solid basis for evaluating privacy protection, differential privacy mechanisms, and robustness enhancement methods. Studies conducted on this dataset can clearly demonstrate the potential advantages of differential-privacy-enhanced federated learning in addressing the dual challenges of data heterogeneity and privacy protection.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies show that introducing differential privacy into distributed modeling can protect user data while improving system reliability in non-ideal environments. This integration not only defends against external attacks but also suppresses interference from malicious clients, thus improving robustness to some extent. However, most existing work still takes a single perspective, often emphasizing either privacy protection or robustness. A systematic framework that optimizes both aspects simultaneously is still lacking. Therefore, exploring differential-privacy-enhanced federated learning to build more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	Acc	Precision	Recall	F1-Score
Fedbiot[14]	83.2	82.5	81.7	82.1
SatFed[15]	84.6	84.0	83.5	83.7

FedAC[16]	85.8	85.1	84.7	84.9
Fedmut[17]	86.3	85.9	85.2	85.5
Ours	89.7	89.2	88.6	88.9

From the comparative experimental results, it can be seen that the proposed method outperforms existing baselines across all core metrics. In terms of accuracy, precision, and recall, the method shows significant advantages. In particular, it achieves an F1-Score of 88.9, which is at least 3.4 percentage points higher than other methods. This indicates that the integration of differential privacy and robustness enhancement mechanisms not only mitigates instability under heterogeneous data distributions but also demonstrates stronger competitiveness in overall performance.

Further analysis shows that traditional federated learning methods often involve trade-offs between privacy and performance. For example, Fedbiot and SatFed protect user data, yet model performance still declines to some extent. In contrast, the proposed method introduces noise clipping and robust aggregation strategies. These strategies effectively compensate for the utility loss caused by differential privacy alone and maintain high predictive performance under privacy constraints. This feature is highly relevant to sensitive data scenarios in real-world applications.

The results also reveal that improved methods, such as FedAC and Fedmut approach the proposed method in some metrics. However, they remain weaker in robustness and generalization. The dual improvements of the proposed method in recall and precision indicate its stronger ability to capture useful features while reducing errors caused by noise or malicious clients. This enables the system to remain stable in complex environments, which is crucial for resisting adversarial attacks and handling abnormal data distributions.

Overall, the experimental results confirm the dual advantages of the proposed differential-privacy-enhanced federated learning framework in both privacy protection and robustness. It breaks through the limitations of prior methods that focus on single-point optimization and achieves a coordinated balance between privacy and performance. Therefore, the method can provide more reliable support for intelligent systems in high-risk domains such as finance, healthcare, and smart cities. It also offers new insights and practical pathways for combining privacy protection with robust AI.

This paper also presents an experiment on the sensitivity of the learning rate to the experimental results, and the experimental results are shown in Figure 2.

From the results shown in the figure, it can be observed that different learning rates have a clear impact on model performance. For accuracy, the model improves as the learning rate increases from 0.001 to 0.01, reaching the best performance at 0.01. When the learning rate is further increased to 0.05, performance decreases. This indicates that a very small learning rate leads to slow convergence and insufficient learning of data features, while an excessively large learning rate may cause oscillations and reduce global convergence stability.

The trend of precision is similar to that of accuracy. This suggests that the adjustment of the learning rate affects not only overall prediction correctness but also the ability to discriminate positive samples. With a moderate learning rate, the model can better maintain stable decision boundaries under the interference of differential privacy noise, which reduces misclassification. This phenomenon shows that a proper learning rate can mitigate the utility loss caused by privacy protection mechanisms and enhance model usability.

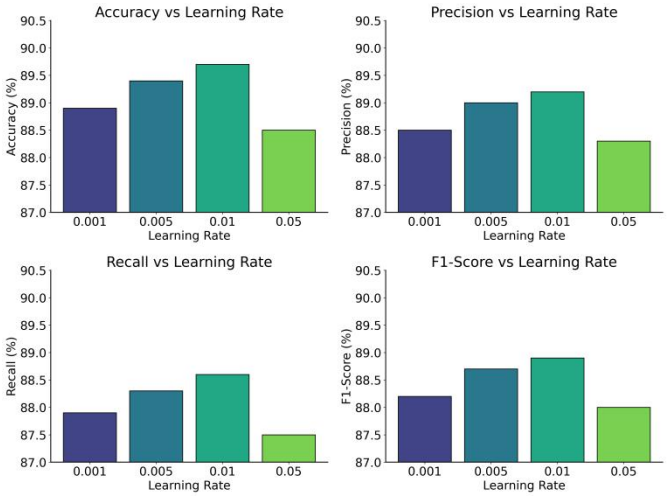

Figure 2. Sensitivity experiment of learning rate to experimental results

For recall, the change with respect to the learning rate also shows a unimodal pattern. When the learning rate is too low, the model captures limited features, leading to insufficient recognition of positive samples. With a moderate learning rate, recall improves significantly, suggesting that the model can more comprehensively cover the target classes. When the learning rate becomes too high, recall decreases sharply, reflecting that feature extraction is disrupted by the combined effect of privacy noise and update oscillations.

F1-Score, as a comprehensive metric, shows a trend consistent with precision and recall. It achieves the best value at a learning rate of 0.01, indicating that the model reaches a relatively optimal balance between accuracy and coverage at this point. This further confirms that under privacy constraints and robustness requirements, the proposed method can maintain stable performance advantages through proper hyperparameter configuration. It also provides practical guidance for deploying differential-privacy-enhanced federated learning in real applications.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies have shown that introducing differential privacy into distributed modeling can protect user data while improving system reliability under non-ideal conditions. This integration can defend against external attacks and suppress interference from malicious clients, thereby improving robustness to some extent. However, most existing work still emphasizes either privacy protection or robustness in isolation. A systematic framework that optimizes both aspects at the same time is still lacking. Therefore, exploring

differential-privacy-enhanced federated learning to build more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

This paper also presents an experiment on the impact of client sampling rate on experimental results, and the experimental results are shown in Figure 3.

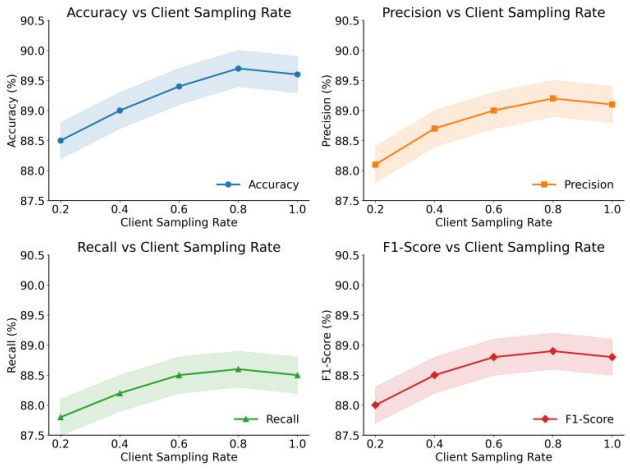

Figure 3. Experiment on the impact of client sampling rate on experimental results

The experimental results show that changes in client sampling rate have a significant impact on model performance. In terms of accuracy, performance steadily improves as the sampling rate increases from 0.2 to 0.8, reaching the best value at 0.8. When the rate increases further to 1.0, accuracy decreases slightly. This suggests that although a higher sampling rate allows the use of more client data, it may also introduce more noise and heterogeneity, which can cause performance fluctuations.

The trend of precision is similar to that of accuracy. At low sampling rates, the model fails to capture sufficient client-specific features, and its ability to distinguish positive samples is limited. With moderate sampling rates, precision improves significantly, showing that the model can better separate positive and negative samples. However, when the rate reaches the maximum, performance improvement becomes negligible or even declines slightly. This further confirms that the sampling rate needs to balance efficiency and stability.

For recall, increasing the sampling rate greatly improves the model's ability to cover positive samples. At low rates, many positive samples are missed, leading to poor recall. As the sampling rate rises, recall steadily increases and peaks near 0.8. This indicates that greater client participation helps the model learn more comprehensive features, thereby improving sensitivity to positive samples. However, when the rate becomes too high, recall decreases slightly, reflecting that data distribution differences and accumulated privacy noise may interfere with model consistency.

The trend of F1-Score combines the performance of precision and recall and shows a similar unimodal curve. The best value is achieved at a sampling rate of 0.8, indicating that the model reaches an optimal balance between accuracy and coverage at this point. This result shows that a reasonable client sampling rate not only finds a trade-off between privacy protection and computational cost but also improves the robustness and practicality of differential-privacy-enhanced federated learning frameworks. It provides valuable guidance for real-world deployment.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies have shown that introducing differential privacy into distributed modeling can protect user data while improving system reliability under non-ideal conditions. This integration can defend against external attacks and suppress interference from malicious clients, thereby improving robustness to some extent. However, most existing work still emphasizes either privacy protection or robustness in isolation. A systematic framework that optimizes both aspects at the same time is still lacking. Therefore, exploring differential-privacy-enhanced federated learning to build more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

This paper further presents an experiment on the sensitivity of client data imbalance to experimental results, and the experimental results are shown in Figure 4.

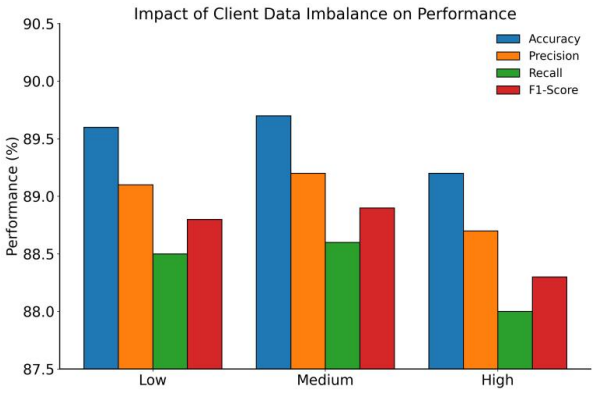

Figure 4. Experiment on the sensitivity of client data imbalance to experimental results

The experimental results show that data imbalance among clients has a clear impact on model performance. Overall, as the imbalance level increases, all four core metrics decrease to varying degrees. When the data distribution is relatively balanced, the model can effectively learn features from different classes and maintain high levels of accuracy and overall performance. In contrast, under highly imbalanced conditions, the scarcity of minority class samples weakens the model's ability to capture the global distribution, leading to performance degradation.

For accuracy and precision, the performance gap between low and moderate imbalance is small. This indicates that the model can still maintain good discriminative ability under moderate distribution differences. However, as the imbalance intensifies, precision declines more sharply, showing that the model is more likely to misclassify minority class samples. This finding reveals that differential-privacy-enhanced

federated learning still requires more robust mechanisms to distinguish positive and negative samples under high imbalance.

The change in recall is particularly evident. Under low and moderate imbalance, recall remains relatively high, suggesting that the model can capture positive samples comprehensively. Under high imbalance, recall drops significantly, reflecting insufficient recognition of minority class samples. This decline highlights the challenge that data distribution poses to model robustness. It also shows that when privacy protection and distribution heterogeneity coexist, the model can be constrained by inadequate coverage of scarce samples.

The overall metric F1-Score follows the same trend as the previous metrics. It remains relatively stable under moderate imbalance but decreases sharply under high imbalance. This indicates that when client data distributions differ too much, the model struggles to maintain a balance between precision and recall. In summary, these results suggest that although differential-privacy-enhanced federated learning can resist the impact of data heterogeneity to some extent, it still faces performance bottlenecks under extreme imbalance. This also implies that future research should further optimize robust aggregation and regularization strategies to mitigate the negative impact of distribution imbalance.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies have shown that introducing differential privacy into distributed modeling can protect user data while improving system reliability under non-ideal conditions. This integration can defend against external attacks and suppress interference from malicious clients, thereby improving robustness to some extent. However, most existing work still emphasizes either privacy protection or robustness in isolation. A systematic framework that optimizes both aspects at the same time is still lacking. Therefore, exploring differential-privacy-enhanced federated learning to build more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

This paper also gives the impact of the anti-noise amplitude on the experimental results, and the experimental results are shown in Figure 5.

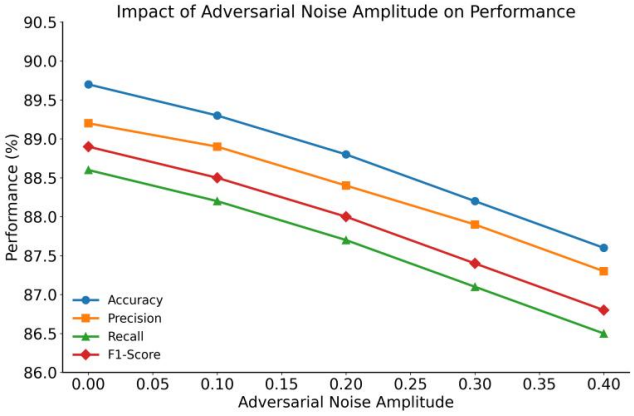

Figure 5. The impact of anti-noise amplitude on experimental results

The experimental results show that increasing the amplitude of adversarial noise significantly affects the overall performance of the model. As the noise amplitude increases, accuracy shows a steady downward trend. This indicates that under strong interference, both the convergence ability and the stability of decision boundaries are damaged. This finding suggests that although differential privacy mechanisms enhance security, the model still shows vulnerability when facing additional adversarial noise. Stronger robustness strategies are required to maintain performance.

The curve of precision shows that larger noise amplitudes reduce the model's ability to discriminate positive samples. This leads to more negative samples being misclassified as positive. When noise is small, the model maintains high precision, which means that in low-disturbance environments, differential-privacy-enhanced federated learning can balance privacy protection and predictive performance. However, when noise becomes too strong, this balance is broken. The model sacrifices accuracy in recognition while maintaining privacy.

For recall, the downward trend is even more evident. As noise amplitude increases, the model's ability to cover positive samples declines. The recognition of minority classes becomes weaker, and missed detections grow more severe. This shows that recall is one of the most sensitive metrics under high uncertainty. It also reflects the limitations of combining differential privacy and robustness mechanisms when strong adversarial interference is present.

The trend of F1-Score is consistent with precision and recall, showing an overall decline. This indicates that the balance between accuracy and coverage is disrupted by noise from both sides. The experimental results demonstrate that although differential-privacy-enhanced federated learning can resist adversarial risks to some extent, a stronger robustness design is still needed when noise amplitude is high. This emphasizes that in studies combining privacy protection and robustness, adaptability to adversarial environments must be considered. Only then can the model ensure usability and trustworthiness in real-world deployment.

Based on this background, the integration of differential privacy and federated learning has gradually become a research focus in recent years. Studies have shown that introducing differential privacy into distributed modeling can protect user data while improving system reliability under non-ideal conditions. This integration can defend against external attacks and suppress interference from malicious clients, thereby improving robustness to some extent. However, most existing work still emphasizes either privacy protection or robustness in isolation. A systematic framework that optimizes both aspects at the same time is still lacking. Therefore, exploring differential-privacy-enhanced federated learning to build more robust AI systems is not only an extension of existing research but also a necessary direction for advancing trustworthy artificial intelligence.

5. Conclusion

This study focuses on a differential-privacy-enhanced federated learning framework, aiming to address the dual challenges of data privacy protection and model robustness.

The proposed method integrates differential privacy mechanisms with robust aggregation strategies. In this way, it ensures user data security while mitigating the negative impact of distribution heterogeneity, adversarial noise, and abnormal client behaviors on model performance. Through systematic comparative and sensitivity experiments, the study demonstrates that the framework achieves superior results across multiple performance metrics, highlighting the value and potential of combining differential privacy with robustness methods.

The results indicate that proper hyperparameter configuration and well-designed privacy budgets are crucial for ensuring system stability and utility. In multidimensional experiments involving learning rate, client sampling rate, and data imbalance, the model shows strong robustness in accuracy, precision, recall, and F1-Score. This not only highlights the algorithmic advantages of the method but also confirms its adaptability to the complexity of real-world applications. These findings further emphasize the importance of differential privacy mechanisms in achieving both privacy compliance and efficient modeling.

From an application perspective, the proposed framework provides new solutions for intelligent systems in sensitive domains such as financial risk control, healthcare, and smart cities. In these scenarios, data are often highly private and strongly heterogeneous, and traditional methods struggle to balance security and effectiveness. By integrating differential privacy with robust design, this study offers strong support for the trustworthiness and usability of models in deployment. It can therefore better serve critical tasks such as risk management, medical diagnosis, and public safety.

In summary, this study contributes theoretical and methodological innovations to the deep integration of differential privacy and federated learning. It also provides empirical evidence for the development of robust artificial intelligence systems. The significance lies not only in enriching the research system of privacy protection and distributed modeling but also in offering practical pathways to address privacy compliance and security risks. In the future, as application scenarios expand and security demands increase, differential-privacy-enhanced federated learning frameworks are expected to show greater value in more complex tasks, driving the advancement of trustworthy artificial intelligence across industries.

References

[1] Gu X, Li M, Xiong L. Precad: Privacy-preserving and robust federated learning via crypto-aided differential privacy[J]. arXiv preprint arXiv:2110.11578, 2021.

[2] Zhu H, Ling Q. Bridging differential privacy and byzantine-robustness via model aggregation[J]. arXiv preprint arXiv:2205.00107, 2022.

[3] Zhang Z, Hu R. Byzantine-robust federated learning with variance reduction and differential privacy[C]//2023 IEEE Conference on Communications and Network Security (CNS). IEEE, 2023: 1-9.

[4] Fu J, Chen Z, Han X. Adap dp-fl: Differentially private federated learning with adaptive noise[C]//2022 IEEE international conference on trust, security and privacy in computing and communications (TrustCom). IEEE, 2022: 656-663.

[5] Feng S, Mohammady M, Hong H, et al. Harmonizing Differential Privacy Mechanisms for Federated Learning: Boosting Accuracy and Convergence[C]//Proceedings of the Fifteenth ACM Conference on Data and Application Security and Privacy. 2024: 60-71.

[6] X. Ren, S. Yang, C. Zhao, J. McCann, and Z. Xu, "Belt and braces: When federated learning meets differential privacy," Communications of the ACM, vol. 67, no. 12, pp. 66-77, 2024.

[7] Guo S, Yang J, Long S, et al. Federated learning with differential privacy via fast Fourier transform for tighter-efficient combining[J]. Scientific Reports, 2024, 14(1): 26770.

[8] Riess A, Ziller A, Kolek S, et al. Complex-Valued Federated Learning with Differential Privacy and MRI Applications[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Cham: Springer Nature Switzerland, 2024: 191-203.

[9] X. Gu, M. Li, and L. Xiong, "PreCAD: Privacy-preserving and robust federated learning via crypto-aided differential privacy," arXiv preprint arXiv:2110.11578, 2021.

[10] Naseri M, Hayes J, De Cristofaro E. Local and central differential privacy for robustness and privacy in federated learning[J]. arXiv preprint arXiv:2009.03561, 2020.

[11] Fu J, Hong Y, Ling X, et al. Differentially private federated learning: A systematic review[J]. arXiv preprint arXiv:2405.08299, 2024.

[12] Qi T, Wang H, Huang Y. Towards the robustness of differentially private federated learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2024, 38(18): 19911-19919.

[13] Lyu L, Yu H, Ma X, et al. Privacy and robustness in federated learning: Attacks and defenses[J]. IEEE transactions on neural networks and learning systems, 2022, 35(7): 8726-8746.

[14] Wu F, Li Z, Li Y, et al. Fedbiot: Llm local fine-tuning in federated learning without full model[C]//Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024: 3345-3355.

[15] S. Li, E. C. H. Ngai, and T. Voigt, "An experimental study of byzantine-robust aggregation schemes in federated learning," IEEE Transactions on Big Data, vol. 10, no. 6, pp. 975-988, 2023.

[16] Zhang Y, Chen H, Lin Z, et al. FedAC: An Adaptive Clustered Federated Learning Framework for Heterogeneous Data[J]. arXiv preprint arXiv:2403.16460, 2024.

[17] Hu M, Cao Y, Li A, et al. FedMut: Generalized federated learning via stochastic mutation[C]//Proceedings of the AAAI conference on artificial intelligence. 2024, 38(11): 12528-12537.