Computers and Artificial Intelligence; Vol.3, No.4; 2026
E-ISSN: 3006-3728 P-ISSN: 3006-371X
DOI: https://doi.org/10.70267/cai.v3n4.178185
Published by: Zeus Press

Differential Privacy and Federated Learning Technologies in Privacy Protection for Medical Big Data

Penghui Zhang*

Silesian College of Intelligent Science and Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China

**Corresponding author: Penghui Zhang.*

Abstract

With the rapid development of big data technology, the value of data in various industries has become increasingly prominent, but the problem of privacy leakage has also received more and more attention. How to protect user privacy while mining the value of data has become an important problem that urgently needs to be solved in the field of big data analysis. This paper reviews and summarizes research related to privacy-preserving big data analysis from three levels: the algorithm level, the framework level, and the application level. First, the application and optimization methods of differential privacy (DP) in deep learning model training are introduced, including key technologies such as the differentially private stochastic gradient descent (DP-SGD) framework, gradient clipping, and privacy budget optimization. Then, the problem of improving model efficiency under the federated learning (FL) framework is discussed, with a focus on communication efficiency optimization and personalized FL methods under non-independent and identically distributed (non-IID) data. Finally, taking medical big data as a typical scenario, the research status and challenges of privacy protection systems are discussed. At the end of the paper, the existing problems in current research and future development directions are also preliminarily prospected.

Keywords

big data, privacy protection, differential privacy, federated learning, medical data

1. Introduction

In the information age, data have become an important factor of production. However, data often contain a large amount of personal sensitive information. If they are not properly protected during collection, storage, and analysis, serious privacy leakage may occur [1]. In recent years, a series of data protection regulations have been issued in China and abroad, such as the General Data Protection Regulation (GDPR) of the European Union and the Personal Information Protection Law of China (PIPL), which have put forward stricter requirements for data use. Against this background, how to conduct effective data analysis without leaking privacy has become a hot issue of common concern in academia and industry. At present, privacy-preserving big data analysis technologies have developed into several branches, among which Differential privacy (DP) and federated learning (FL) are two relatively representative directions [2-3].

DP injects carefully designed random noise into data or algorithms so that attackers cannot infer individual information from query results. With its strict mathematical definition and provable privacy guarantee, it has been widely applied in the field of data publication and analysis [4]. As a distributed machine learning

© 2026 The Author(s). Authors retain copyright in their articles.

paradigm, FL allows multiple participants to collaboratively build models without sharing raw data, thereby reducing the risk of data leakage from the source [2].

This paper systematically reviews privacy-preserving big data analysis technologies from three levels. At the algorithm level, it focuses on the application and optimization of DP in deep learning. At the framework level, it discusses efficiency improvement strategies in FL. At the application level, it takes medical big data as an example to analyze the practical status of privacy protection systems, and it looks ahead to the challenges and future development directions.

2. Application Optimization of Differential Privacy in Deep Learning

2.1 Differential Privacy Basics and the DP-SGD Framework

DP is a privacy protection model with a strict mathematical definition. It realizes privacy protection by adding an appropriate amount of noise to query or analysis results. Formally, if the change in the output distribution of an algorithm is limited after a single record is added to or deleted from the input data, the algorithm is said to satisfy DP [4]. Here, the idea that the "change is not large" is controlled by the privacy budgets ε and δ. The smaller ε is, the higher the degree of privacy protection is, but data availability may decline, so a trade-off between the two is required.

When DP is applied to deep learning, the most classical method is DP-SGD. Its core idea is to protect the privacy of training data through gradient clipping and noise injection [5]. The basic idea of DP-SGD is as follows: on the basis of standard stochastic gradient descent (SGD), the gradient of each sample is first clipped to limit its norm so that it does not exceed a preset threshold; then Gaussian noise is injected into the clipped gradients; finally, the model parameters are updated with the noisy gradients. This mechanism constrains the influence of a single sample on model updates, thereby realizing privacy protection.

The principle of DP-SGD is relatively intuitive, but accurate tracking of the privacy budget is its key challenge. The privacy budget estimation obtained by the advanced composition theorem used in early studies is relatively conservative. As an important variant of DP, Renyi differential privacy (RDP) uses Renyi divergence to measure privacy loss and can track the privacy budget more tightly. It has become a standard analysis tool in DP deep learning [6].

In addition to DP-SGD, there are also methods such as the Private Aggregation of Teacher Ensembles (PATE) framework. PATE adopts a teacher-student architecture. It first trains multiple teacher models with private data and then provides labels for the student model through voting results. In addition, the combination of DP and generative adversarial networks (GANs) is also an important direction. By imposing DP constraints on the gradients of the discriminator, high-quality synthetic data can be generated while protecting the privacy of training data. Methods such as dynamic gradient clipping further improve the usability of this type of model [7]. Overall, because DP-SGD is simple to implement and widely applicable, it remains the current mainstream method for DP deep learning.

2.2 Gradient Clipping and Privacy Budget Optimization

In DP-SGD, gradient clipping is a key step. Its function is to limit the magnitude of the gradient of each sample, thereby controlling the scale of noise that needs to be injected. However, the selection of the clipping threshold is a core challenge. If the threshold is too large, the required noise increases and model utility decreases. If the threshold is too small, excessive loss of gradient information occurs [5].

Traditional DP-SGD uses a fixed clipping threshold, which is often difficult to optimize in practice. For this reason, researchers have proposed various adaptive clipping strategies. For example, Zhou and Qian proposed the FSDP algorithm, which dynamically adjusts the clipping threshold according to the gradient distribution [8]. Zhang et al. also proposed a dynamic decay method for the clipping threshold, which gradually reduces the clipping threshold as training proceeds [9].

In their review, Sha et al. systematically summarized the optimization methods of DP-SGD into three levels: the gradient level, the iteration level, and the architecture level [5]. The gradient level mainly involves improvements in clipping and noise addition methods. The iteration level involves the adjustment of

© 2026 The Author(s). Authors retain copyright in their articles.

hyperparameters such as the learning rate and batch size. The architecture level designs models that are more suitable for DP training from the perspective of the network structure itself.

In recent years, the rise of large language models (LLMs) has also brought new opportunities for DP. Some studies have found that large models may obtain a better privacy-utility trade-off under DP training. This may be because the parameter space of large models is sufficiently large, so the influence of added noise is relatively small. Lu et al. combined DP with low-rank adaptation (LoRA) and proposed the DP-LoRA method, which is specially used to protect sensitive information in large model training [10].

2.3 Integrated Extension of Differential Privacy and Federated Learning

Research in China on the intersection of DP and deep learning is relatively active. Cao et al. proposed a DP FL algorithm based on the functional mechanism. By perturbing the objective function rather than directly perturbing gradients, it effectively avoids the accuracy loss caused by gradient clipping and also performs well in non-IID data scenarios [11].

The review article by Sha et al. also comes from this research team. They divided the optimization methods of DP-SGD into three levels and established a relatively systematic classification framework [5]. Zhang et al. proposed FedBDP, a personalized FL method based on Bregman divergence and DP, which can both protect privacy and adapt to differences in data distribution among different clients [12].

Wang et al. studied the relatively new direction of federated unlearning. They proposed the FedUR mechanism and combined it with DP technology to realize the “right to be forgotten” [13]. This direction has strong practical significance, because current privacy regulations include requirements for the right to deletion.

Overall, domestic research on DP deep learning has gradually moved from following to keeping pace, and in some subdivided directions it even shows a tendency to lead.

3. Efficiency Improvement Methods under the Federated Learning Framework

3.1 Federated Learning Basics and the FedAvg Algorithm

The core concept of FL can be summarized as “data stay still while the model moves”: each participant uses its own local data to train a model, uploads model parameters to a central server for aggregation, and then receives the global model from the server for continued iterative training [2]. In this process, raw data always remain local, which fundamentally reduces the risk of privacy leakage.

FedAvg is the most classical algorithm in FL. It aggregates the model parameters of each client through weighted averaging [2]. Its process is as follows: each client performs multiple rounds of SGD updates locally and then uploads parameters; the server performs weighted aggregation according to the proportion of data held by each client.

Although the principle of FedAvg is simple, it faces multiple challenges in practical deployment. The first is the problem of communication efficiency. In large-scale FL scenarios, frequent communication between numerous clients and the server generates huge overhead. Gu and Bai systematically summarized the security and privacy issues of FL models in their review [3]. The second major problem is data heterogeneity. The data distributions of different clients may vary greatly. This non-IID situation can cause the problem of “client drift,” which seriously affects the convergence speed and final performance of the model. Existing reviews have systematically summarized this issue [14].

3.2 Communication Efficiency Optimization Methods

Communication efficiency is a key issue in the implementation and application of FL. Zheng et al. systematically classified FL communication efficiency optimization methods into four categories: client selection, model compression, topology reconstruction, and training strategy optimization [15].

The core idea of client selection strategies is to selectively invite some high-quality clients to participate in each round of training. This not only reduces communication overhead, but also prevents straggling clients

© 2026 The Author(s). Authors retain copyright in their articles.

from slowing down the training process. Liu et al. further studied participant selection from the perspective of privacy protection and used secure multi-party computation (SMPC) to realize secure top-k client search [16].

Model compression is another important direction. Common methods include quantization, sparsification, and low-rank decomposition. However, the degree of compression must be reasonably controlled to avoid significantly damaging model performance. Training strategy optimization reduces communication rounds by adjusting hyperparameters such as the number of local training epochs and the learning rate. For example, the SCAFFOLD algorithm uses control variates to correct client drift, while FedProx adds a proximal term to the local loss function to restrict the difference between the local model and the global model. These are typical methods for addressing statistical heterogeneity [14].

3.3 Personalized Federated Learning under Non-IID Data

Data heterogeneity is a very difficult problem in FL. If the data distributions of different clients vary greatly, the trained global model may perform poorly on all sides. To solve this problem, personalized FL has become a research hotspot in recent years. The basic idea of personalized FL is not to pursue a global model that is the same for all clients, but to allow each client to have a personalized model suitable for its own data distribution [13].

The method system of personalized FL is relatively rich. Regularization-based methods allow each client to learn the global model and the local model at the same time, and constrain the difference between the two through divergence or regularization terms [13]. The second category is based on model splitting. The third category is based on knowledge distillation. Ping and Fu proposed a personalized FL method for dual skewness of concept and label distributions [17]. In addition, there are clustering-based methods, such as the Fourier personalized FL proposed by Jin and Chen [18].

However, most existing personalized FL methods have only been verified on public datasets, and their implementation practice for industrial scenarios is still relatively limited.

4. Privacy Protection Systems for Medical Big Data

4.1 Privacy Protection Needs and Challenges of Medical Data

Medical data are one of the most sensitive types of data and are directly related to personal health conditions and privacy. Once medical data are leaked, patients may suffer discrimination in employment, insurance, and other aspects, with very serious consequences. Therefore, the medical field has particularly strict requirements for privacy protection.

At the same time, medical data also contain great value. Large-scale medical data analysis helps reveal the laws of disease occurrence and development, improve diagnostic accuracy, and promote new drug research and development. In recent years, FL technology has made significant research progress in medical image diagnosis and other directions [19].

However, medical data sharing faces multiple difficulties. Laws and regulations restrict medical institutions from sharing patient data arbitrarily; data are dispersed among institutions and form “data islands”; and data types such as electronic health records (EHRs), medical images, and genomics are diverse, inconsistent in format, and uneven in quality.

Zhang et al. studied a multi-keyword range search scheme for privacy protection of medical big data. They used searchable encryption technology to realize efficient retrieval while ensuring data confidentiality [20]. Relevant reviews point out that a single privacy protection technology is difficult to meet the diversified needs of medical scenarios, and the integration of multiple technologies is a future development trend [21].

4.2 Application of Federated Learning in Medicine

FL exactly meets the needs of the medical scenario in which “data do not leave the hospital and models are built collaboratively”: hospitals train models locally and only exchange model parameters, while raw data always remain inside the hospital [19].

© 2026 The Author(s). Authors retain copyright in their articles.

FL has broad application prospects in the medical field. In medical imaging, FL can be used to collaboratively train image classification, segmentation, and detection models. Existing reviews have systematically summarized FL technology in medical imaging, covering research progress in classification, segmentation, and various other tasks [19].

For EHRs, Zhang and Tan proposed an FL privacy protection method for multimodal medical data, which can process image data and medical record data at the same time [22]. Wang et al. proposed the homomorphic encryption-based federated learning privacy-preserving scheme (HEFLPS) architecture, which uses homomorphic encryption (HE) to protect the transmission security of model parameters [23]. However, the computational overhead of HE is relatively large, and how to improve efficiency is a problem that needs to be solved.

The review by Nie et al. pointed out that the implementation of FL in the medical field still faces multiple challenges, such as data heterogeneity, privacy compliance, and incentive mechanisms [24]. In particular, the problem of uneven data quality is prominent. Data labeling standards may differ among hospitals, which can seriously affect the performance of FL models.

4.3 Multiple Privacy Protection Paths for Medical Data

DP also has important applications in medical data protection. Unlike FL, DP focuses more on adding noise to the results of data analysis so that attackers cannot infer individual information from the results. In medical scenarios, DP can be used in multiple links, such as data publication, statistical queries, and model training.

Some studies have systematically analyzed the algorithms and performance of DP FL and investigated changes in model convergence and accuracy under different noise levels and different data distributions. The PPVFL method proposed by Li et al. is aimed at vertical FL. It uses a max-min strategy to protect privacy and has application prospects in medical diagnosis and other fields [25].

There are also many domestic studies on medical privacy protection systems. Wang and Kan studied access control in mobile health (mHealth) and proposed a fine-grained policy-hiding and traceable decentralized access control scheme [26]. Han et al. studied medical data access control and secure sharing in public health events. By combining blockchain and attribute-based access control (ABAC), they realized traceable data sharing [27].

Overall, domestic research on medical privacy protection still mainly focuses on theoretical and algorithmic exploration, and implementation systems for large-scale clinical applications remain relatively lacking.

5. Challenges and Prospects

In summary, privacy-preserving big data analysis technologies have developed rapidly, but they still face many challenges that urgently need to be solved.

First, there is the balance between privacy protection and data utility. Whether DP or FL is used, privacy protection comes with a certain cost. How to minimize the impact on data utility and system efficiency while ensuring sufficient privacy protection is a long-standing core problem. Especially in fields such as medicine that have very high accuracy requirements, the accuracy loss caused by privacy protection may be difficult to tolerate.

Second, there is the deep integration of multiple privacy protection technologies. Current research mostly focuses on a single technology, while actual scenarios often require the collaborative application of multiple technologies. Technology integration is not a simple superposition, and the compatibility and cumulative effects among different technologies must be fully considered.

Third, there is engineering implementation from algorithms to systems. At present, many privacy protection algorithms remain at the stage of theoretical analysis or experiments on small-scale datasets, and cases truly deployed in large-scale production systems are limited. This involves both technical challenges and nontechnical factors such as standards and regulations.

© 2026 The Author(s). Authors retain copyright in their articles.

Fourth, there are new privacy protection challenges in the era of large models. The rapid development of LLMs has brought new privacy risks, such as how to efficiently implement DP in large-scale model training and how to prevent models from memorizing sensitive information in training data.

Looking to the future, privacy-preserving big data analysis will show development trends of technology integration, vertical application, standardization, and intelligent methods.

6. Conclusion

This paper systematically reviews privacy-preserving big data analysis technologies from three dimensions: DP, FL, and medical big data applications. As a privacy protection method with strict theoretical guarantees, DP is being increasingly widely applied in deep learning. FL provides an effective distributed solution to the data island problem, and it still has broad research space in efficiency optimization and personalized modeling. As a typical application scenario, medical big data puts forward more special and stricter requirements for privacy protection.

In summary, privacy-preserving big data analysis is a dynamic research field. With the continuous evolution of big data and artificial intelligence technologies, the importance of privacy protection will become increasingly prominent.

References

[1] Tan, Z., & Zhang, L. (2020). Survey on privacy-preserving machine learning. Journal of Software, 31(7), 2127-2156.

[2] Liu, Y., Chen, H., Liu, Y., & Li, C. (2022). Privacy-preserving techniques in federated learning. Journal of Software, 33(3), 1057-1092.

[3] Gu, Y., & Bai, Y. (2023). Research progress on model security and privacy of federated learning. Journal of Software, 34(6), 2833-2864.

[4] Zhang, X., & Meng, X. (2014). Differential privacy in data publication and analysis. Chinese Journal of Computers, 37(4), 927-949.

[5] Sha, H., Sun, L., Liu, Y., Xue, D., Wu, Y., Li, C., & Chen, H. (2026). Optimization of differentially private stochastic gradient descent: Methods, theories and applications. Chinese Journal of Computers, 49(6), 1307-1341.

[6] Wang, K., Wang, Y., Liu, J., & Deng, J. (2025). Graph convolutional collaborative filtering recommendation algorithm based on Renyi differential privacy. Journal of Software, 36(3), 1202-1217.

[7] Chen, S., Cai, J., & Sun, L. (2023). Differential privacy generative adversarial network algorithm with dynamic gradient threshold clipping. Journal of Computer Applications, 43(7), 2065-2072.

[8] Zhou, Z., & Qian, X. (2022). A differential privacy protection algorithm for deep neural networks. Journal of Electronics and Information Technology, 44(5), 1773-1781.

[9] Zhang, X., Zhang, L., & Zhang, Z. (2025). Federated learning method under user-level local differential privacy. Journal of Computer Research and Development, 62(2), 472-487.

[10] Lu, Y., Jiang, Y., Feng, H., Guo, Z., & Xiang, S. (2025). Differential low-rank adaptation-based sensitive information protection for large language model training. Computer Engineering, 51(10), 1-14. https://doi.org/10.19678/j.issn.1000-3428.00252845.

[11] Cao, S., Chen, C., Tang, P., & Su, S. (2023). Differentially private federated learning with functional mechanism. Chinese Journal of Computers, 46(10), 2178-2195.

[12] Zhang, S., Zhang, J., Zhu, G., Long, S., & Li, Z. (2024). Personalized federated learning method based on Bregman divergence and differential privacy. Journal of Software, 35(11), 5249-5262. https://doi.org/10.13328/j.cnki.jos.007032.

© 2026 The Author(s). Authors retain copyright in their articles.

[13] Wang, T., Zhai, L., Yu, Y., Yang, T., Zhang, X., & Ren, X. (2025). Differentially private federated unlearning mechanism based on update residual. Scientia Sinica Informationis, 55(11), 2704-2721.

[14]Yu, H., Fan, J., Sun, Y., Dong, H., & Xi, E. (2025). Survey of statistical heterogeneity in federated learning. Journal of Computer Applications, 45(9), 2737-2746.

[15] Zheng, J., Liu, B., Lin, W., & Xie, J. (2025). Survey of communication efficiency for federated learning. Computer Science, 52(2), 1-7.

[16] Liu, T., Yang, A., Weng, J., Chen, M., Liu, Y., & Zeng, H. (2026). Privacy-preserving heterogeneous federated learning participant selection based on secure multi-party computation. Journal of Software. Advance online publication. https://doi.org/10.13328/j.cnki.jos.007622.

[17] Ping, F., & Fu, X. (2026). Personalized federated learning for concept and label distribution drift. Computer Science, 53(7), 433-441.

[18] Jin, T., & Chen, S. (2023). Spectral clustering-based Fourier personalized federated learning. Journal of Electronics and Information Technology, 45(6), 1981-1989.

[19] Liu, Y., Dai, Y., & Chen, G. (2025). Review of federated learning in medical image processing. Computer Science, 52(1), 183-193.

[20] Zhang, M., Huang, J., & Han, L. (2021). Multi-keyword range search scheme for privacy protection of medical big data. Journal of Software, 32(10), 3266-3282.

[21] Li, F., Li, H., Niu, B., & Qiu, W. (2022). Academic connotation and research trends of privacy computing. Chinese Journal of Network and Information Security, 8(6), 1-8.

[22] Zhang, L., & Tan, Z. (2023). Federated learning privacy protection method for multimodal medical data. Computer Science, 50(S2), 933-940.

[23] Wang, B., Li, H., Wang, J., & Guo, Y. (2023). Privacy-preserving federated learning architecture for medical data. Journal of Xidian University, 50(5), 166-177.

[24]Nie, W., Xin, L., & Li, Z. (2022). Research on application of federated learning in medical informatization. Journal of Medical Informatics, 43(10), 12-16.

[25] Li, R., Liu, T., Zheng, H., Chen, J., Liu, Z., & Ji, S. (2024). Privacy preservation method for vertical federated learning based on max-min strategy. Acta Automatica Sinica, 50(7), 1373-1388. https://doi.org/10.16383/j.aas.c211233.

[26] Wang, J., & Kan, H. (2024). Fine-grained policy-hiding and traceable decentralized access control scheme in mHealth. Journal of Computer Research and Development, 61(6), 1525-1535.

[27] Han, G., Wang, J., Luo, W., & Lu, Y. (2023). Research on medical data access control and secure sharing in public health events. Journal of Cyber Security, 8(1), 40-54.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

© 2026 The Author(s). Authors retain copyright in their articles.

Copyrights

Copyright for this article is retained by the author (s), with first publication rights granted to the journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/4.0/).

© 2026 The Author(s). Authors retain copyright in their articles.