Federated Learning with Differential Privacy: Decreasing Noise Mechanism

Aisha Sartaj*
aishasartaj1@ucla.edu

Ssuyung Yeh*
ssuyung@ucla.edu

Vijayasree Garapati*
vijayasree@g.ucla.edu

Yen-Yen Kuo*
angelakuo0214@g.ucla.edu

Abstract

Federated learning, combined with differential privacy, offers a robust approach to safeguarding sensitive data during distributed model training. In this study, we propose a decreasing noise mechanism to optimize the balance between privacy and model accuracy. By dynamically adjusting the gradient clipping size through a noise control parameter, *noise_gamma*, we explore the trade-offs between convergence rate and accuracy under fixed privacy budgets (ϵ, δ). Experiments conducted using Gaussian and Laplace noise on the MNIST dataset demonstrate that decreasing noise improves model accuracy compared to fixed noise scales. Notably, Gaussian noise showed a slight accuracy enhancement for specific *noise_gamma* values, while Laplace noise yielded a modest 2–4% improvement. Our findings underline the importance of parameter tuning in achieving privacy-utility trade-offs, paving the way for future explorations into more complex datasets and models.

1 Introduction

Data privacy is a critical concern across individual, organizational, and societal domains. In distributed large-scale machine learning, preserving the confidentiality of sensitive data is vital to ensure trust and compliance with privacy regulations. Federated learning and differential privacy have emerged as essential tools for safeguarding data in such settings.

Federated Learning enables decentralized model training by keeping data localized on devices, transmitting only aggregated model parameters rather than raw data [1,2,3]. This decentralized framework is particularly significant in sectors like healthcare, finance, and law, where data security is paramount [4,5]. However, despite its advantages, federated learning is not immune to privacy risks, as gradients can inadvertently reveal sensitive information.

To address this, Differential Privacy (DP) provides a mathematically rigorous framework to protect individual data contributions. It ensures that the inclusion or exclusion of any individual's data has a negligible impact on the model's output, shielding participants from additional risks [6]. In traditional DP implementations, noise is consistently added to gradients during the training process. However, this approach does not adapt to the evolving nature of training dynamics. As gradients diminish over time, constant noise levels can result in suboptimal privacy-utility tradeoffs, potentially degrading model performance.

Recent strategies aim to mitigate these issues by dynamically adjusting the noise mechanism [7,8]. Techniques such as progressively reducing noise, scaling noise based on gradient reduction, and using gradient clipping to control noise magnitude have been proposed. These approaches seek to find the optimal balance between privacy preservation and model performance. In this work, we propose and evaluate a decreasing noise mechanism that adjusts noise dynamically, leveraging hyperparameters such as the noise level and clipping threshold. Our experiments assess the impact of these strategies on model accuracy and convergence rate, using the sensitivity of these hyperparameters as a key focus.

Striking a balance between privacy and utility remains a complex challenge. Through this study, we aim to provide insights into optimizing this tradeoff, paving the way for more efficient and privacy-preserving federated learning frameworks.

2 Related Work

2.1 Federated Learning

Federated learning (FL) is regarded as a promising framework to enable model training across distributed devices or institutions while preserving data privacy. By keeping data localized and transmitting only aggregated model updates, FL addresses key privacy concerns associated with centralized data storage. FL has been used in plenty applications in privacy-sensitive fields such as healthcare and mobile devices, where the sharing of raw data is often restricted. Lu et al. [4] introduces a federated learning framework for computational pathology on gigapixel whole slide images (WSIs), enabling collaborative model training across institutions while preserving privacy. Using weakly-supervised multiple instance learning (MIL), the model leverages slide-level annotations for cancer classification and survival prediction without requiring pixel-level labels.

Building on this, Wei et al. [9] explore this in a multi-institutional framework, proposing a federated learning model that adapts noise levels according to each institution's data characteristics to optimize privacy and model convergence. These approaches highlight federated learning's potential to advance medical research by maintaining rigorous privacy standards while allowing broader data utility.

2.2 Differential Privacy

Machine learning often relies on large datasets containing sensitive personal information, raising privacy concerns if the data is compromised. To address this, privacy-preserving techniques, such as differential privacy, are implemented to safeguard confidential data while balancing model utility [6]. Banse et al. [1] introduced a refined noise-adding process that adjusts noise levels based on model layers and sensitivity to preserve essential features while limiting privacy risk.

Additionally, Yousefpou et al. [13] introduces Opacus, a PyTorch library designed for differential privacy. Opacus optimizes the process by computing gradients per sample in batches, simplifying deep learning model training with differential privacy to as few as two lines of code. In addition, the model employs attention-based pooling for interpretability, identifying prognostic regions that aid in patient risk stratification for survival analysis.

Another important technique for implementing differential privacy is the Sparse Vector Technique (SVT). SVT allows specific query answers without significantly consuming the privacy budget, providing efficient privacy maintenance for repetitive queries in federated learning environments. This technique has been adapted in various interactive and non-interactive settings, offering a method to selectively apply privacy controls depending on query outcomes. Recent studies, such as those by Lyu et al. [14], have further optimized SVT, improving its utility in privacy-sensitive applications while avoiding some misapplications seen in prior versions.

Furthermore, optimal noise-adding mechanisms play a critical role in balancing utility and privacy. Research by Geng and Viswanath [7, 8] has indicated the trade-offs of (ϵ, δ)-differential privacy and proposed optimized noise distributions, including the innovative staircase mechanism. By minimizing the noise magnitude and power, this method optimizes the noise-adding process. It outperforms the traditional Laplacian mechanism, particularly in low-privacy settings. These advances highlight the importance of precise mechanism selection and optimization when designing models that need to preserve privacy.

2.3 Applications

Li et al. [2] summarized various applications of federated learning in fields such as healthcare, mobile devices, and industrial engineering. However, T. Li et al. [3] highlights several significant challenges in federated learning, including statistical heterogeneity, high communication costs, system heterogeneity, and privacy issues. Addressing these challenges requires advancements such as heterogeneity diagnostics, extreme communication schemes, novel models for asynchrony, and granular privacy constraints. In healthcare, federated learning with differential privacy has facilitated data collaboration across hospitals without exposing patient data.

Shah et al. [15] discuss federated learning's role in medical imaging, enabling institutions to share model parameters rather than data, which is particularly valuable in settings constrained by privacy regulations. Further discusses the need for heterogeneity diagnostics, extreme communication schemes, novel models for asynchrony and granular privacy constraints. Bonawitz et al. [16] describes a scalable system designed for mobile devices that supports federated learning, with a focus on adapting federated learning to manage a range of diverse tasks. Additionally, Lee and Clifton [17] explored the challenge of selecting the privacy parameter epsilon in differential privacy, noting its dependence on data attributes and query types. It underscores the difficulty in balancing privacy protection with the risk of individual identification.

3 Method

This paper introduces a federated learning framework with differential privacy, incorporating a novel decreasing noise mechanism. In this framework, gradients are clipped on each local device before applying noise, such as Gaussian or Laplacian noise, to ensure privacy. Figure 1 illustrates the workflow, where gradient clipping limits the sensitivity of updates, and noise addition enforces differential privacy guarantees.

Notably, gradients tend to diminish over time during training. This diminishing trend can result in suboptimal privacy-utility tradeoffs if a constant noise level is applied. To address this limitation, we propose an adaptive noise mechanism that reduces the magnitude of added noise dynamically as training progresses. This approach aims to preserve privacy while improving model performance by minimizing unnecessary noise interference in later stages of training.

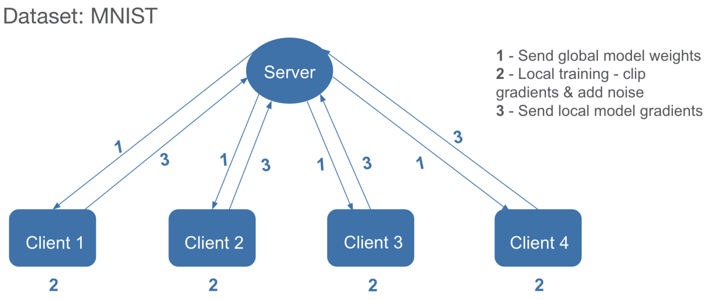
Dataset: MNIST
Server
1 - Send global model weights
2 - Local training - clip gradients & add noise
3 - Send local model gradients
1
3
1
3
1
3
1
3
Client 1
Client 2
Client 3
Client 4
2
2
2
2

Figure 1: Federated learning workflow with gradient clipping and noise addition to ensure differential privacy. The decreasing noise mechanism dynamically adjusts noise levels as training progresses.

Differential privacy [4] is defined as a mechanism that ensures the indistinguishability of outputs for two adjacent datasets D_i and D_i', differing by exactly one record. Formally, it satisfies (ϵ, δ)-differential privacy if for any subset of outputs S , a deterministic function f satisfies:

$$P[f(D_i) \in S] \leq e^{\epsilon} P[f(D_i') \in S] + \delta \tag{1}$$

Here, ϵ and δ represent the privacy budget, where smaller values indicate stronger privacy guarantees.

In the context of gradient descent, the standard deviation σ of Gaussian noise can be computed as:

$$\sigma = \frac{C\sqrt{2\ln(1.25/\delta)}}{\epsilon} \tag{2}$$

where C is the gradient clipping size. In contrast, Laplace noise naturally enforces ϵ-differential privacy with a noise scale b given by:

$$b = \frac{C}{\epsilon} \tag{3}$$

Both mechanisms ensure privacy by bounding the sensitivity of gradients via clipping. Thus we reduce the clip size C gradually during training to lower the noise scale while maintaining the fixed (ϵ, δ) - differential privacy guarantees.

4 Experiment

4.1 Experimental Setup

The proposed approach was evaluated using the MNIST dataset, chosen for its simplicity and widespread use in benchmarking machine learning models. A three-layer Convolutional Neural Network (CNN) was implemented for both worker and master nodes. The CNN consisted of two convolutional layers followed by a fully connected layer, with ReLU activations and max-pooling to process image data efficiently. This architecture serves as a robust baseline for testing privacy-preserving techniques.

The experiment included four clients, each initialized with a gradient clipping size of 0.1. Gaussian noise was added using a delta parameter of 0.00001, ensuring consistency with Laplace noise, which inherently enforces stricter privacy without a delta parameter. The adaptive noise mechanism was applied to dynamically scale noise levels, leveraging the diminishing gradients observed during training.

As explained in the related work section, the privacy budget is defined by epsilon (ϵ) and delta (δ) and determines the level of privacy preservation. The gradient clipping size C will control the noise scale with fixed ϵ and δ. Therefore, in our experiment we can preserve a given privacy budget (ϵ, δ) while reducing the noise scale by decreasing the gradient clipping size C.

4.2 Decreasing noise method

To dynamically reduce the noise scale during training, we set up another hyper parameter *noise_gamma* to control the rate at which the gradient clipping size C decreases in each local epoch. Formally, the clipping size is updated as $C = C \times$ *noise_gamma* after every local epoch. To gradually decrease the clipping size, the value of *noise_gamma* is chosen to be between 0 and 1. A smaller *noise_gamma* would lead to a slower convergence rate since the gradient clipping is stricter. However, with a lower *noise_gamma*, the noise scale would also be decreased by a larger factor after each local epoch and theoretically resulting in a higher accuracy. Therefore, how much the *noise_gamma* should be becomes a trade-off between convergence rate and model accuracy. In our experiments, we analyze the impact of different *noise_gamma* values to find out the optimal settings that enhance model performance while maintaining a fixed privacy budget (ϵ, δ).

4.3 Experimental Design

To evaluate the effectiveness of *noise_gamma*, we conduct two sets of experiments with several noise addition strategies:

1. **Fixed noise version:** Apply noise with various fixed noise scales but different ϵ to the gradients.

2. **Decreasing noise version:** Apply noise using different values of *noise_gamma* to control the gradient noise scale.

In the first experiment, we first test whether a larger noise scale (i.e., a smaller epsilon in differential privacy) actually leads to a lower convergence accuracy and slower convergence rate. In the second experiment, we evaluate the impact of our decreasing noise technique on model accuracy, and analyzing how different values of *noise_gamma* affect to different convergence rate and accuracy.

5 Results

5.1 Fixed noise

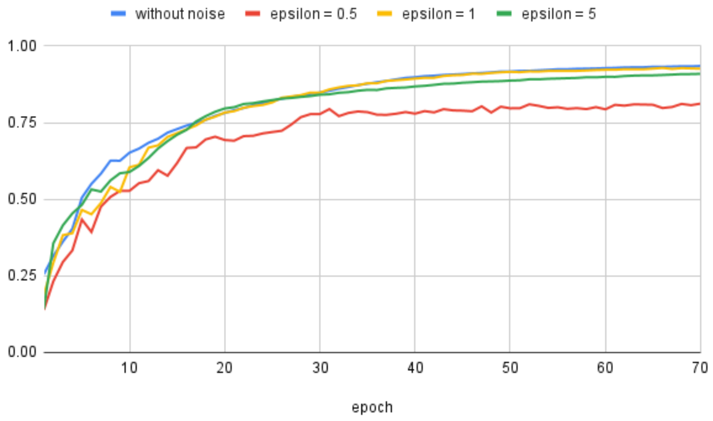

Figure 2: Gaussian without decreasing noise

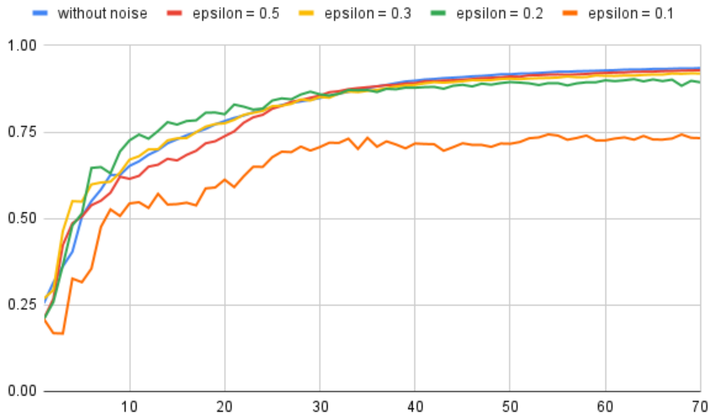

Figure 3: Laplace without decreasing noise

Based on the results shown in Figures 2 and 3, it is clear that lower ϵ correspond to lower model accuracy. This trend is consistent for both Gaussian and Laplace noise without decreasing noise. As ϵ decreases, the noise scale increases. This leads to slower convergence and reduced accuracy. For example, in the Gaussian noise case, models with $\epsilon = 5$ achieve higher final accuracy than model with $\epsilon = 1$ or $\epsilon = 0.5$. Similarly, for Laplace noise, the accuracy decreases significantly when when ϵ drops below 0.5. This demonstrates that smaller privacy budgets result in a trade-off between privacy and utility.

5.2 Decreasing noise

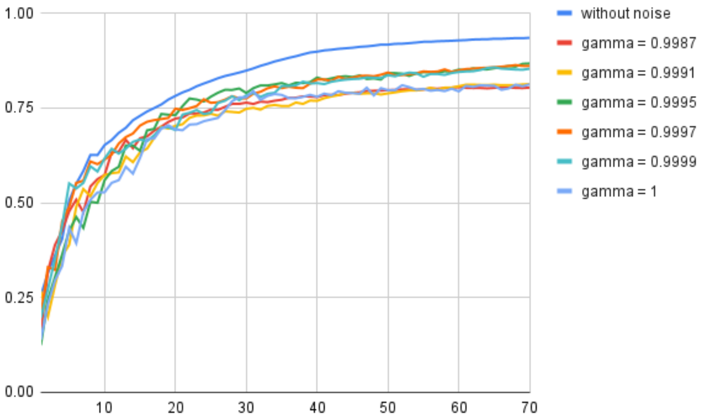

Figure 4: Gaussian with decreasing noise, $\epsilon = 0.5$

Figure 4 shows that the use of the Gaussian noise decreasing technique improves the accuracy of the final model for certain ranges of *noise_gamma*. Specifically, when *noise_gamma* are between 0.9995 and 0.9999, the final accuracy increases slightly compared to the fixed noise model. However, the differences between the accuracies within this range aren't noticeable. When *noise_gamma* is decreased to 0.9991 and 0.9987, the accuracy drops to the same as the model without decreasing noise. This phenomenon is probably due to the significant reduced clipping threshold, which decreases the convergence rate even though the noise is highly reduced.

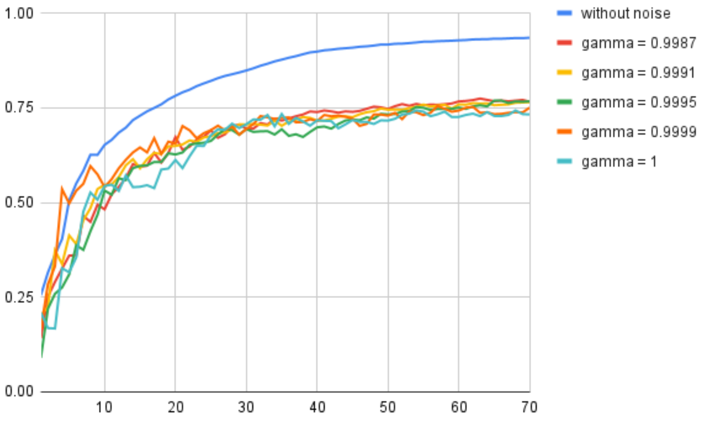

Figure 5: Laplace with decreasing noise, $\epsilon = 0.1$

As shown on Figure 5, the results show that applying the decreasing noise technique with Laplace noise also improve the accuracy. As the value of *noise_gamma* decreases, the accuracy slightly increases. This confirms the effectiveness of reducing noise. However, the improvement is not significant, with the accuracy increasing by only 2-4% compared to the model without decreasing noise. Additionally, the differences between various *noise_gamma* are minimal. This indicates that the impact of decreasing noise on performance remains limited under these settings.

6 Future Works

The proposed framework can be further validated by evaluating its performance on diverse datasets, including those with complex data types such as medical images or time-series data. Future work could explore optimizing the balance between clip size and privacy budget, specifically examining the relationship between the noise scale (gamma) and accuracy improvements. Identifying an optimal balance, referred to as a “sweet spot” could significantly enhance accuracy. Additionally, the approach can be expanded by integrating different model architectures and employing model pruning [10,11] techniques at local devices. Finally, experiments could be conducted to assess the framework’s efficacy when incorporating personalized federated learning strategies [12].

7 Conclusion

This research highlights the efficacy of a decreasing noise mechanism in federated learning with differential privacy. By dynamically scaling noise based on gradient clipping, the proposed approach enhances accuracy while adhering to strict privacy constraints. Gaussian and Laplace noise both benefit from this mechanism, although their sensitivity to *noise_gamma* varies. Notably, while Gaussian noise achieves optimal accuracy within a narrow range of *noise_gamma*, Laplace noise demonstrates more consistent improvements. Despite these advances, the observed gains remain modest, indicating a need for further optimization. This adaptive approach offers a promising pathway toward achieving both high performance and robust privacy in distributed machine learning systems.

8 Contribution

Idea - equal contributions by each team member
Implementation - Aisha Sartaj:30 % (laplace noise and its experiments), Ssuyung Yeh:30% (gaussian noise and its experiments), Vijayasree Garapati:20% (federated learning setup and integration with different noise mechanisms), Yen Yen Kuo:20% (differential privacy setup and fine tuning experiments)
Presentation - equal contributions by each team member
Write up - Aisha Sartaj:20% (abstract, literature review , references), Ssuyung Yeh:20% (literature review, conclusion, future work), Vijayasree Garapati:30% (introduction, literature review , method) Yen Yen Kuo:30% (literature review, experiments, results)

References

[1] Banse, Adrien, and Jan Kreischer. "Federated Learning with Differential Privacy." arXiv preprint arXiv:2402.02230 (2024).

[2] Li, Li, et al. "A review of applications in federated learning." Computers & Industrial Engineering 149 (2020): 106854.

[3] T. Li, A. K. Sahu, A. Talwalkar and V. Smith, "Federated Learning: Challenges, Methods, and Future Directions," in IEEE Signal Processing Magazine, vol. 37, no. 3, pp. 50-60, May 2020, doi: 10.1109/MSP.2020.2975749. keywords: Distributed databases;Data models;Training data;Data privacy;Privacy;Predictive models;Machine learning

[4] Lu, Ming Y., et al. "Federated learning for computational pathology on gigapixel whole slide images." Medical image analysis 76 (2022): 102298.

[5] Truex, Stacey, et al. "A hybrid approach to privacy-preserving federated learning." Proceedings of the 12th ACM workshop on artificial intelligence and security. 2019.

[6] Dwork, Cynthia. "Differential privacy." International colloquium on automata, languages, and programming. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[7] Q. Geng and P. Viswanath, "The Optimal Noise-Adding Mechanism in Differential Privacy," in IEEE Transactions on Information Theory, vol. 62, no. 2, pp. 925-951, Feb. 2016, doi: 10.1109/TIT.2015.2504967. keywords: Privacy;Data privacy;Sensitivity;Databases;Laplace

equations;Probability distribution;Context;Data privacy;randomized algorithm;Data privacy;randomized algorithm

[8] Q. Geng and P. Viswanath, "Optimal Noise Adding Mechanisms for Approximate Differential Privacy," in IEEE Transactions on Information Theory, vol. 62, no. 2, pp. 952-969, Feb. 2016, doi: 10.1109/TIT.2015.2504972. keywords: Privacy;Sensitivity;Data privacy;Cost function;Databases;Laplace equations;Standards;Data privacy;randomized algorithm;Data privacy;randomized algorithm

[9] Wei, Kang, et al. "Federated learning with differential privacy: Algorithms and performance analysis." IEEE transactions on information forensics and security 15 (2020): 3454-3469.

[10] Y. Jiang et al., "Model Pruning Enables Efficient Federated Learning on Edge Devices," in IEEE Transactions on Neural Networks and Learning Systems, vol. 34, no. 12, pp. 10374-10386, Dec. 2023, doi: 10.1109/TNNLS.2022.3166101. keywords: Training;Computational modeling;Data models;Adaptation models;Collaborative work;Servers;Distributed databases;Efficient training;federated learning (FL);model pruning

[11] Jiang, Yuang, et al. "Model pruning enables efficient federated learning on edge devices." IEEE Transactions on Neural Networks and Learning Systems 34.12 (2022): 10374-10386.

[12] S. Vahidian, M. Morafah and B. Lin, "Personalized Federated Learning by Structured and Unstructured Pruning under Data Heterogeneity," 2021 IEEE 41st International Conference on Distributed Computing Systems Workshops (ICDCSW), Washington, DC, USA, 2021, pp. 27-34, doi: 10.1109/ICDCSW53096.2021.00012. keywords: Training;Deep learning;Conferences;Distributed databases;Benchmark testing;Collaborative work;Data models;Deep learning;federated learning;personalization;subnetwork

[13] Yousefpour, Ashkan, et al. "Opacus: User-friendly differential privacy library in PyTorch." arXiv preprint arXiv:2109.12298 (2021).

[14] Lyu, Min, Dong Su, and Ninghui Li. "Understanding the sparse vector technique for differential privacy." arXiv preprint arXiv:1603.01699 (2016)

[15] U. Shah, I. Dave, J. Malde, J. Mehta and S. Kodeboyina, "Maintaining Privacy in Medical Imaging with Federated Learning, Deep Learning, Differential Privacy, and Encrypted Computation," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-6, doi: 10.1109/I2CT51068.2021.9417997. keywords: Training;Deep learning;Differential privacy;Law;Lakes;Regulation;Cryptography;Deep learning;Differential privacy;Federated learning;Medical data;Secure and Privacy-preserving AI

[16] Bonawitz, Keith. "Towards federated learning at scale: Syste m design." arXiv preprint arXiv:1902.01046 (2019).

[17] Jaewoo Lee and Chris Clifton, "How much is enough? choosing ϵ for differential privacy" in Information Security: 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings*, Springer Berlin Heidelberg, 2011.

[18] Geyer, Robin C., Tassilo Klein, and Moin Nabi. "Differentially private federated learning: A client level perspective." arXiv preprint arXiv:1712.07557 (2017).

[19] R. Kumari, D. K. Sah, S. Gupta, K. Cengiz and N. Ivković, "Advancing Medical Recommendations With Federated Learning on Decentralized Data: A Roadmap for Implementation," in IEEE Transactions on Consumer Electronics, vol. 70, no. 1, pp. 2666-2674, Feb. 2024, doi: 10.1109/TCE.2023.3334159. keywords: Data models;Medical diagnostic imaging;Medical services;Federated learning;Distributed databases;Data privacy;Training;Federated learning;personalized medical recommendations;decentralized data;model architecture;and sensitivity analysis