KoalaMamba: A Federated Method with Vision Mamba UNet for Privacy-Preserving Medical Segmentation

Haocheng Kan , Mingpei Cao , Yuesheng Zhu
School of Electronic and Computer Engineering, Peking University, Shenzhen, China
{kanhorst, caomingpei}@alumni.pku.edu.cn, zhuys@pku.edu.cn

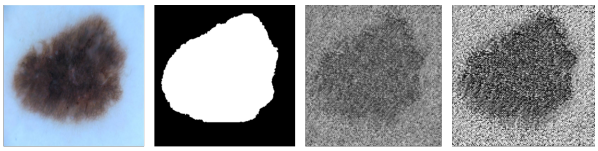
***Abstract*—Distributed machine learning holds strong promise for medical image segmentation in clinical diagnosis and treatment planning. However, collaborative training across multiple medical centers is fundamentally constrained by strict privacy requirements, which limits data sharing and hinders large-scale deployment. This paper presents KoalaMamba, a practical federated learning (FL) method designed for privacy-preserving medical segmentation while remaining compatible with multiple FL optimizers. KoalaMamba integrates client-side differential privacy (DP) into federated training, making it statistically difficult to infer whether any individual record participated in learning, thereby mitigating privacy leakage risks and protecting institutional confidentiality. To improve segmentation accuracy and generalizability, KoalaMamba incorporates Vision Mamba UNet-based architectures, including VMUNet, HVMUNet, and VMUNetV2, as configurable backbones. On the ISIC 2018 dermoscopic segmentation benchmark, KoalaMamba achieves competitive performance across diverse FL settings. In particular, under the SCAFFOLD Non-IID configuration, KoalaMamba is evaluated with DP enabled using the Moment Accountant (MA) at $\epsilon = 10$ and $\delta = 10^{-5}$, reaching 0.7822 IoU and 0.8778 DSC. This DP-constrained result substantially outperforms the UNet baseline and remains competitive with strong Mamba backbones evaluated without DP. Moreover, in separate privacy evaluation experiments, applying DP with a tighter budget ($\epsilon = 1$) reduces the attack success rate (ASR) from 0.92 to 0.19 and decreases the membership inference AUC from 0.89 to 0.51, demonstrating that KoalaMamba effectively mitigates privacy leakage while preserving practical segmentation utility.**

***Index Terms*—Federated Learning, Privacy-Preserving, Medical Image Segmentation.**

I. Introduction

Accurate segmentation of medical images is of fundamental importance in clinical practice. It enables the delineation of anatomical structures such as tumors, organs, and lesions from imaging modalities including computed tomography and magnetic resonance imaging [1], [2]. Automated segmentation systems alleviate the diagnostic burden on radiologists and provide quantitative support for surgical planning and disease progression monitoring. Nevertheless, the development and deployment of robust segmentation models in the healthcare domain remain constrained by three critical factors: the fragmentation of data across institutions, the need for strict privacy protection, and the lack of adaptable methods capable of generalizing across heterogeneous data distributions.

Centralized learning paradigms that aggregate raw patient data onto a single server present a significant privacy risk, potentially violating privacy regulations such as the Health Insurance Portability and Accountability Act[1] and the General Data Protection Regulation[2]. Furthermore, they are vulnerable to privacy attacks, such as model inversion, which reconstruct sensitive inputs from model parameters. A representative example is shown in **Fig. 1**, where inversion attacks applied to dermoscopic segmentation models expose visual features of skin lesions. Differential privacy (DP) has been widely adopted to counter such threats by introducing calibrated noise during training to statistically obscure the presence of any individual sample [3]–[5].



(a) (b) (c) (d)

Fig. 1. Visual comparison of model inversion attack [3] outcomes on ISIC 2018 images. (a) Original dermoscopic image. (b) Segmentation mask. (c) Reconstruction with differential privacy. (d) Reconstruction without differential privacy. KoalaMamba with DP significantly reduces the recoverability of sensitive image features, mitigating inversion attack effectiveness.

To address the aforementioned limitations, we present **KoalaMamba**, a modular federated learning method for privacy-preserving medical image segmentation. The proposed method integrates three complementary components: (1) flexible support for multiple federated learning strategies including FedAvg [6], FedProx [7], and SCAFFOLD [8]; (2) an architecture family based on visual state space modeling, encompassing VMUNet and HVMUNet [9], [10]; and (3) differential privacy mechanisms realized through noise injection and gradient clipping, with privacy accounting performed via Rényi Differential Privacy and Moment Accountant [11], [12].

The main contributions of this study are as follows:

- We propose KoalaMamba, a unified federated segmentation method that jointly addresses Non-IID data heterogeneity and client-side differential privacy, providing a

[1]See HIPAA guidelines at https://www.cdc.gov/phlp/php/resources/health-insurance-portability-and-accountability-act-of-1996-hipaa.html

[2]See GDPR information at https://gdpr-info.eu/

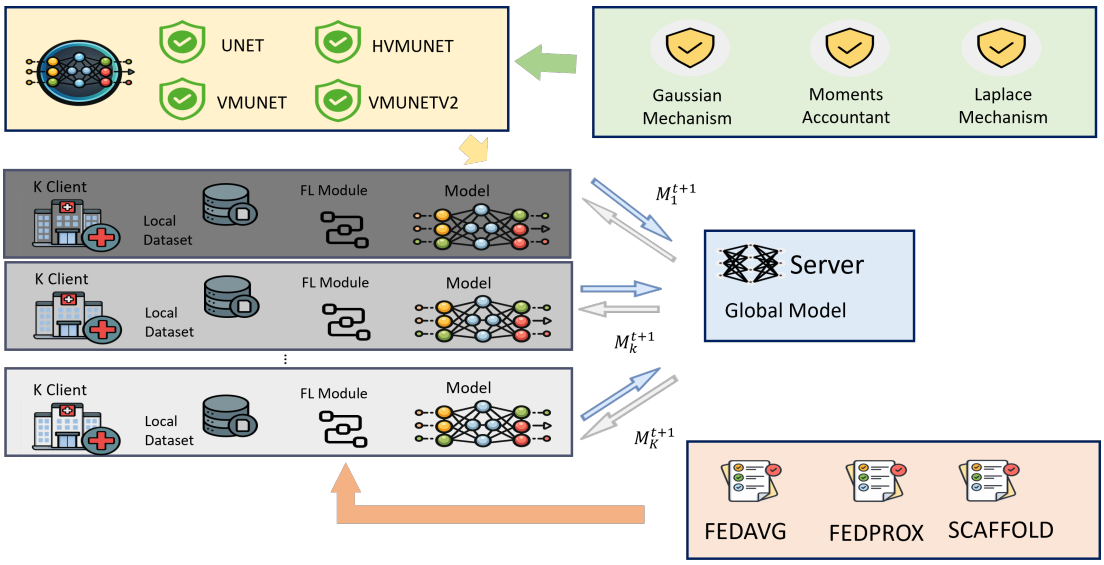


Fig. 2. KoalaMamba's architecture integrates federated learning strategies, segmentation backbones, and differential privacy mechanisms. The modular design enables flexible combinations, such as HVMUNet with FedProx and Moment Accountant, for comparative evaluation in clinical scenarios.

practical training protocol for privacy-sensitive medical image segmentation.

- We design a configurable client update mechanism with explicit privacy accounting, supporting Gaussian, Laplace, and Moment Accountant schemes under a unified formulation, enabling systematic analysis of privacy and utility trade-offs in federated settings.
- Under explicit privacy constraints, KoalaMamba consistently achieves competitive segmentation performance across multiple federated optimizers and Non-IID scenarios, demonstrating strong robustness while preserving formal privacy guarantees for real-world clinical deployment.

II. Related Work

A. Medical Image Segmentation

Recent advances in neural architectures have substantially improved segmentation accuracy. UNet, with its encoder-decoder topology and skip connections, has been widely adopted in biomedical image analysis [1], [13], [14]. Its successors, such as transformer-based models [15], further exploit long-range dependencies to enhance representational capacity. More recently, Mamba-based models have gained prominence by replacing attention mechanisms with structured state space representations that allow efficient sequence modeling while reducing computational overhead [16], [17].

VMUNet incorporates the visual state space mechanism to capture global and local features, and HVMUNet introduces high-order scanning to improve granularity and reduce feature redundancy [9], [10]. These models have demonstrated superior performance in complex segmentation tasks, especially under data-limited conditions.

B. Federated Learning

Federated learning has emerged as an effective alternative to centralized training, particularly in sensitive domains such as healthcare [18], [19]. However, FL suffers from practical limitations including statistical heterogeneity across clients, convergence instability, and excessive communication overhead [20], [21]. FedAvg performs global aggregation by averaging local updates, but its effectiveness degrades under Non-IID distributions [6]. FedProx mitigates this by regularizing local updates, while SCAFFOLD incorporates control variates to address client drift [7], [8].

In the context of medical image segmentation, federated learning has been paired with advanced backbone models such as VMUNet to achieve performance levels comparable to centralized methods. Studies have demonstrated that these configurations can preserve data privacy while ensuring segmentation accuracy across institutions with distinct data characteristics [22].

C. Privacy Protection in FL

Despite these advances, federated learning remains susceptible to privacy attacks such as membership inference and model inversion. Differential privacy has therefore become a critical component of secure FL systems. Common implementations involve adding noise to model updates and applying gradient clipping to limit sensitivity. Techniques such as Gaussian and Laplace mechanisms are widely used, while Moment Accountant and Rényi Differential Privacy provide rigorous tools for privacy budget tracking and composition analysis [23], [24].

Algorithm 1 KoalaMamba: Server-Side Process

Input: Initial Model M_0, Clients $\{C_1, \ldots, C_N\}$, Rounds T
Output: Final Global Model M_T
for $t = 1$ to T **do**
  $\mathcal{K}_t \subseteq$ Random($\{C_1, \ldots, C_N\}$)
  for each $k \in \mathcal{K}_t$ **in parallel do**
    $M_k^{t+1} \leftarrow$ Update(M_t, k)
  end for
  $M_{t+1} \leftarrow$ Aggregate($\{M_k^{t+1}\}_{k \in \mathcal{K}_t}$)
end for
return M_T

Algorithm 2 KoalaMamba: Client-Side Process with Configurable FL and DP

Input: Global Model M_t, Local Data $\mathcal{D}_k$, $\mathcal{F} \in \{$FedAvg, FedProx, SCAFFOLD$\}$: FL Method, $\mathcal{P} \in \{\mathcal{G}, \mathcal{L}, \mathcal{G}_{MA}\}$: DP Mechanism
Initialize $M_k \leftarrow M_t$
for $e = 1$ to E **do**
  for each batch $b \in \mathcal{D}_k$ **do**
    $\nabla \ell \leftarrow$ Grad(M_k, b) *# Compute gradient*
    $\nabla \ell \leftarrow \nabla \ell / \max(1, \|\nabla \ell\|_2 / G)$ *# Clip to bound sensitivity*
    $M_k \leftarrow M_k - I(\eta \nabla \ell, \mathcal{F})$ *# Optimize based on* $\mathcal{F}$
  end for
end for
if $\mathcal{P} \equiv \mathcal{G}$ **then**
  $\xi \sim \mathcal{N}(0, \sigma^2 I)$ *# Gaussian Noise*
else if $\mathcal{P} \equiv \mathcal{L}$ **then**
  $\xi \sim$ Laplace(λ) *# Laplace Noise*
else if $\mathcal{P} \equiv \mathcal{G}_{MA}$ **then**
  $\xi \sim \mathcal{N}(0, \sigma^2 I)$ *# Gaussian Noise (MA accounting)*
end if
return $M_k^{t+1} = M_k + \xi$

III. Methodology

A. Overview

Federated learning (FL) has shown strong potential in privacy-sensitive domains such as medical image analysis. However, it faces key challenges in mitigating privacy leakage, achieving convergence under heterogeneous (Non-IID) data, and maintaining communication efficiency. To address these, we propose KoalaMamba, a modular and privacy-preserving FL method tailored for robust segmentation. It integrates three core components: federated learning strategies such as FedAvg, FedProx and SCAFFOLD; segmentation backbones exemplified by VMUNet and HVMUNet; and differential privacy mechanisms incorporating Rényi DP and the Moment Accountant for secure local training, as illustrated in **Fig. 2**.

Throughout this paper, we follow standard terminology, using "IID" to refer to independent and identically distributed client data and "Non-IID" to describe heterogeneous distributions. KoalaMamba enables flexible configuration of FL algorithms and segmentation models, which can be independently selected rather than executed concurrently. This design facilitates empirical benchmarking under consistent privacy conditions while maintaining adaptability to deployment requirements.

B. Federated Learning Method

KoalaMamba follows a modular client-server paradigm, where a global model M on the server is iteratively updated through client-side training. In each round, selected clients receive M_t, train locally, and return M_k^{t+1} for aggregation. This process repeats for T rounds. To reflect data heterogeneity, KoalaMamba supports both IID and Non-IID settings; the former uses uniform partitioning, while the latter employs label-skew or shard-based schemes to simulate real-world imbalance. The method's core logic is presented in Algorithm 1 and Algorithm 2, detailing server aggregation and client updates with configurable privacy. KoalaMamba supports FedAvg, FedProx, and SCAFFOLD for optimization, and integrates DP via Gaussian, Laplace, or MA noise injection, with RDP-based accounting for MA. These components support flexible deployment under diverse privacy and data conditions.

KoalaMamba incorporates a suite of Federated Learning (FL) strategies to facilitate distributed model training across diverse data landscapes. This method supports the **FedAvg** [6], aggregating client contributions via a weighted average of their local model updates. To better handle statistical heterogeneity inherent in decentralized datasets, KoalaMamba also supports **FedProx** [7], introducing a proximal term during local training to stabilize convergence and mitigate divergence from the global model. Furthermore, this method includes **SCAFFOLD** [8], which utilizes control variates to reduce variance across client updates to improve convergence speed and accuracy.

For the segmentation model architecture, KoalaMamba offers compatibility with various backbones. This includes the widely adopted **UNet**, a standard benchmark especially prevalent in biomedical image segmentation [1]. Additionally, it integrates modern Mamba-based architectures, including **VMUNet**, which enhanced long-sequence modeling capabilities [9], and **HVMUNet**, an extension that incorporates high-order spatial interactions to capture more complex patterns [10]

By default, KoalaMamba uses weighted FedAvg for server-side aggregation. When differential privacy is applied, each client uploads a noise-perturbed model, and the server performs aggregation without modifying the rule. The global update follows $M_{t+1} = \sum_{k \in \mathcal{K}_t} \beta_k (M_k^{t+1} + \xi_k)$, where β_k reflects local data size and ξ_k is noise from the selected DP mechanism.

C. Differential Privacy in Federated Learning

KoalaMamba incorporates differential privacy (DP) on the client side to prevent information leakage during model aggregation. It supports multiple DP mechanisms, including Gaussian noise, Laplace noise, and the Gaussian Moment Accountant (MA), to flexibly adapt to various privacy-utility requirements in federated learning scenarios.

1) Gaussian Mechanism with DP-SGD. Each client downloads the global model M_G^t from the server at round t and initializes its local model:

$$M_k^t = M_G^t \tag{1}$$

After local training for E epochs on local dataset $\mathcal{D}_k$, the client obtains an updated model M_k^{t+1}. Before uploading to

TABLE I
IOU RESULTS OF MODELS ON ISIC 2018 (KOALAMAMBA WITH DP).

IoU	Centralized	FedAvg IID	FedAvg NonIID	FedProx IID	FedProx NonIID	Scaffold IID	Scaffold NonIID
UNet	**0.7192**	0.5669	0.5466	0.6093	0.5731	0.5961	0.5652
VMUNet	0.7920	0.7691	0.7840	0.7743	0.7826	0.7902	**0.7885**
HVMUNet	**0.7990**	0.7854	0.7952	0.7139	0.7107	0.7958	0.7978
VMUNetV2	**0.7812**	0.7258	0.6503	0.7091	0.7004	0.7112	0.6443
Ours	0.6953	0.7179	0.7040	**0.7189**	0.6932	**0.7840**	0.7822

TABLE II
DSC RESULTS OF MODELS ON ISIC 2018 (KOALAMAMBA WITH DP).

DSC	Centralized	FedAvg IID	FedAvg NonIID	FedProx IID	FedProx NonIID	Scaffold IID	Scaffold NonIID
UNet	**0.8367**	0.7236	0.7068	0.7573	0.7286	0.7469	0.7222
VMUNet	0.8839	0.8695	0.8789	0.8728	0.8781	0.8828	**0.8817**
HVMUNet	**0.8883**	0.8798	0.8859	0.8331	0.8309	0.8863	0.8875
VMUNetV2	**0.8771**	0.8411	0.7881	0.8298	0.8238	0.8313	0.7836
Ours	0.8203	0.8358	0.8263	**0.8365**	0.8188	**0.8789**	0.8778

TABLE III
AVERAGE SEGMENTATION PERFORMANCE (IOU AND DSC) UNDER DIFFERENT PRIVACY MECHANISMS AND PRIVACY BUDGETS (ϵ). NOTE: GAU. = GAUSSIAN, LAP. = LAPLACE, MA = MOMENT ACCOUNTANT

ϵ	Avg IoU			Avg DSC		
	Gau.	Lap.	MA	Gau.	Lap.	MA
1.0	0.0608	0.4029	0.4841	0.1075	0.6035	0.6618
5.0	0.6866	0.6814	0.6911	0.8101	0.8073	0.8126
10.0	0.6948	0.6851	0.7012	0.8136	0.8089	0.8197
20.0	0.7051	0.6953	0.7067	0.8243	0.8136	0.8255
30.0	0.7082	0.6972	0.7072	0.8272	0.8161	0.8263

the server, Gaussian noise $\xi \sim \mathcal{N}(0, \sigma^2\mathbb{I})$ is added to ensure (ϵ, δ)-DP:

$$\overline{M}_k^{t+1} = M_k^{t+1} + \xi \tag{2}$$

The server aggregates these noisy models using weighted averaging:

$$M_G^{t+1} = \sum_k \beta_k \overline{M}_k^{t+1} \tag{3}$$

where $\beta_k = \frac{n_k}{\sum_j n_j}$ denotes the fraction of training data held by client k, ensuring fairness in update contribution.

2) Laplace Mechanism with Gradient Clipping. Alternatively, clients can opt to update gradients rather than model weights. Each client's per-batch gradient $\mathbf{u}_k$ is clipped to a norm bound B to bound sensitivity:

$$\mathbf{u}_k = \frac{\mathbf{u}_k}{\max(1, \|\mathbf{u}_k\|/B)} \tag{4}$$

Gradients from selected clients are averaged:

$$\mathbf{u} = \frac{1}{|\mathcal{K}|} \sum_k \mathbf{u}_k \tag{5}$$

The updated model $\mathbf{m}_c$ is computed with learning rate γ as:

$$\mathbf{m}_c = \mathbf{m}_s - \gamma \mathbf{u} \tag{6}$$

Laplace noise is injected for privacy protection:

$$\widetilde{\mathbf{m}}_c = \mathbf{m}_c + \mathrm{Lap}\left(\frac{\Delta\phi}{\epsilon}\right) \tag{7}$$

where $\Delta\phi$ represents the ℓ_1-sensitivity of the clipped gradient update. Equation (2) represents the Gaussian mechanism, while Equation (7) reflects Laplace-based differential privacy.

3) Privacy Accounting with RDP and MA. To track cumulative privacy cost over multiple training rounds, KoalaMamba adopts Rényi Differential Privacy (RDP), which extends (ϵ, δ)-DP using Rényi divergence of order α and enables tighter composition bounds. For Gaussian noise, we implement the Moment Accountant (MA), an efficient RDP-based method for tracking the accumulated privacy loss. After training, RDP guarantees are converted back into (ϵ, δ) form for interpretability. We fix $\delta = 10^{-5}$ throughout our experiments, and vary ϵ to assess the privacy-utility trade-off. Notably, smaller ϵ implies stronger privacy with higher noise, while larger ϵ favors utility by reducing noise, consistent with standard DP theory. This accounting method ensures that KoalaMamba offers both quantifiable and enforceable privacy guarantees across multi-round federated training.

IV. EVALUATION

Datasets and Experimental Setup. We evaluate KoalaMamba on the ISIC 2018 dataset [2], [25], a widely used benchmark in dermoscopic lesion segmentation. The dataset contains more than 2,500 images with pixel-level annotations. To simulate federated training, the dataset is partitioned into multiple client subsets representing different institutions. Both IID and Non-IID splits are considered: IID splits are obtained through random partitioning, while Non-IID splits are constructed by shard-based label skew to emulate heterogeneous client distributions.

KoalaMamba supports three backbone models: UNet (baseline), VMUNet, and HVMUNet. Federated training is performed under three strategies (FedAvg, FedProx, SCAFFOLD)

TABLE IV
PRIVACY AND UTILITY EVALUATION UNDER DIFFERENT DP MECHANISMS ON ISIC 2018 WITH VMUNET.

Mechanism	ϵ	SSIM ↑	PSNR (dB) ↑	IoU ↑	DSC ↑	ASR ↓	AUC (MIA) ↓
w/o DP (baseline)	–	0.851	27.4	0.8012	0.8898	0.92	0.89
Gaussian	10	0.703	22.8	0.6948	0.8136	0.74	0.71
	5	0.621	20.5	0.6866	0.8101	0.63	0.60
	1	0.392	13.9	0.0608	0.1075	0.19	0.51
Laplace	10	0.678	21.5	0.6851	0.8089	0.69	0.67
	5	0.572	19.6	0.6814	0.8073	0.58	0.58
	1	0.341	13.2	0.4029	0.6035	0.16	0.47
MA	10	0.712	23.0	0.7012	0.8197	0.63	0.65
	5	0.668	21.8	0.6911	0.8126	0.52	0.57
	1	0.481	17.3	0.4841	0.6618	0.22	0.49

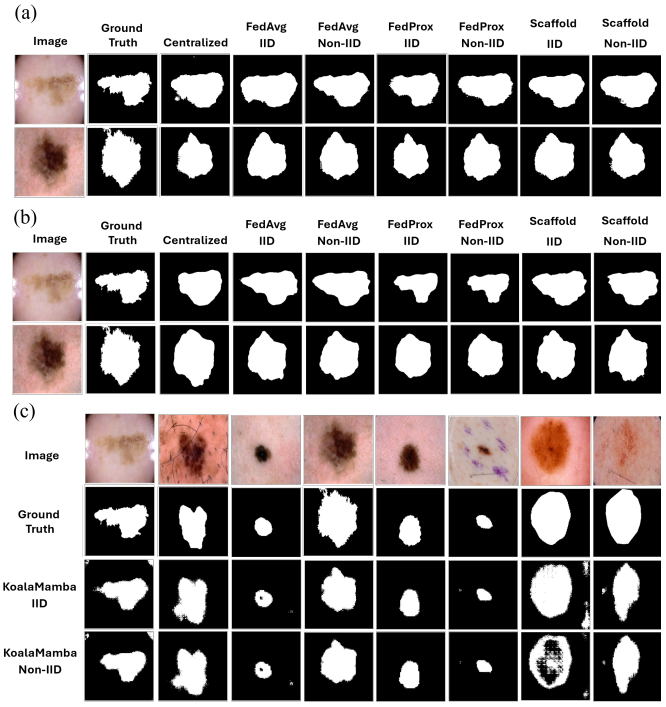


Fig. 3. Qualitative and quantitative comparisons on the ISIC 2018 dataset. (a) Results of centralized training, FedAvg, FedProx, and SCAFFOLD using the VMUNet model. (b) Results of centralized training, FedAvg, FedProx, and SCAFFOLD using the HVMUNet model. (c) Results of the proposed KoalaMamba framework, combining VMUNet with FedAvg and differential privacy mechanisms.

for T global rounds, with E local epochs per client. Differential privacy (DP) is applied client-side using Gaussian, Laplace, or Moment Accountant (MA) mechanisms. For each privacy mechanism, we vary the budget $\epsilon \in \{1, 5, 10, 20, 30\}$ at a fixed $\delta = 10^{-5}$. All models are trained with a combined Dice and cross-entropy loss. Unless otherwise specified, the reported results of KoalaMamba (Ours) in Tables I and II are obtained with client-side differential privacy enabled using the Moment Accountant (MA). We fix the privacy parameter $\delta = 10^{-5}$ and set $\epsilon = 10$ to represent a practical privacy and utility trade-off. All non-KoalaMamba baselines are evaluated without differential privacy to characterize the utility upper bound under each federated learning configuration.

Evaluation Metrics. Segmentation performance is assessed using Intersection over Union (IoU) and Dice Similarity Coefficient (DSC):

$$\text{IoU} = \frac{|P \cap G|}{|P \cup G|}, \text{DSC} = \frac{2|P \cap G|}{|P| + |G|}.$$

where P is the predicted segmentation and G is the ground truth. For privacy assessment, we employ metrics from inversion and membership inference attacks: the attack success rate (ASR) and the area under the ROC curve (AUC). In addition, we measure image-level similarity with Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR) to quantify reconstruction fidelity under different privacy regimes.

Baselines and Results. We compare KoalaMamba against centralized training and conventional federated baselines. Tables I and II report IoU and DSC across UNet, VMUNet, HVMUNet, VMUNetV2, and KoalaMamba. The results show that while centralized training achieves the highest performance, KoalaMamba remains competitive under federated settings. Notably, under the SCAFFOLD Non-IID setting, KoalaMamba is evaluated with client-side DP enabled (MA, $\epsilon = 10$, $\delta = 10^{-5}$) and attains 0.7822 IoU and 0.8778 DSC. While strong Mamba backbones without DP still achieve higher absolute accuracy, KoalaMamba substantially outperforms the UNet baseline and demonstrates competitive performance under explicit privacy constraints.

Table III summarizes average segmentation performance across DP mechanisms. At tight privacy budgets ($\epsilon = 1$), all mechanisms degrade utility; however, MA consistently preserves higher IoU/DSC compared to Gaussian and Laplace. Table IV further demonstrates that DP reduces ASR from 0.92 to as low as 0.19 and MIA AUC from 0.89 to 0.51, confirming effective privacy protection.

Fig. 3 visualizes training dynamics and segmentation outcomes. While centralized models remain upper bounds, KoalaMamba balances performance and privacy under federated settings, making it a practical candidate for clinical deployment. **Within our evaluated settings, KoalaMamba maintains competitive segmentation performance under an explicit differential privacy budget, relative to non-DP Mamba-based baselines.**

V. Discussion

KoalaMamba achieves robust segmentation under both IID and Non-IID settings while providing formal differential privacy guarantees. In the VMUNet-FedAvg setting, it reaches 0.7179 IoU and 0.8358 DSC under IID, with comparable results under Non-IID conditions. Although its Non-IID DSC is slightly lower than VMUNet and HVMUNet, it still outperforms UNet and remains close to centralized training. Table IV further shows that differential privacy is effective: when $\epsilon = 1$, ASR drops from 0.92 to 0.19 and MIA AUC decreases from 0.89 to 0.51. Among the tested mechanisms, Moment Accountant provides the best privacy-utility balance, preserving segmentation quality at moderate privacy budgets ($\epsilon \geq 10$). These results confirm KoalaMamba as a flexible and privacy-preserving framework for federated medical segmentation, although adversarial threats such as model poisoning still need to be addressed in future work.

VI. Conclusion

KoalaMamba is a federated learning framework for privacy-sensitive medical image segmentation. It integrates multiple FL strategies, advanced segmentation models, and differential privacy to balance utility and confidentiality. Experiments on ISIC 2018 demonstrate strong performance under both IID and Non-IID settings, close to centralized baselines. Privacy results further show reduced attack success rates and MIA leakage, especially at $\epsilon = 1$, while the Moment Accountant achieves the best trade-off between segmentation quality and protection. These findings confirm its practical value for secure collaborative deployment in decentralized clinical environments.

References

[1] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 2015, pp. 234–241.

[2] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," *arXiv preprint arXiv:1902.03368*, 2019. [Online]. Available: https://arxiv.org/abs/1902.03368

[3] N. Subbanna, M. Wilms, A. Tuladhar, and N. D. Forkert, "An analysis of the vulnerability of two common deep learning-based medical image segmentation techniques to model inversion attacks," *Sensors*, vol. 21, no. 11, p. 3874, 2021.

[4] Q. Li, Y. Diao, Q. Chen, and B. He, "Federated learning on non-iid data silos: An experimental study," in *2022 IEEE 38th international conference on data engineering (ICDE)*. IEEE, 2022, pp. 965–978.

[5] E. Bagdasaryan and P. Kairouz, "Towards sparse federated analytics: Location heatmaps under distributed differential privacy with secure aggregation," in *PoPETs/PETS-Privacy Enhancing Technologies Symposium*, 2022.

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Artificial intelligence and statistics*. PMLR, 2017, pp. 1273–1282.

[7] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, and V. Smith, "Federated optimization in heterogeneous networks," *Proceedings of Machine learning and systems*, vol. 2, pp. 429–450, 2020.

[8] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*. PMLR, 2020, pp. 5132–5143.

[9] J. Ruan and S. Xiang, "Vm-unet: Vision mamba unet for medical image segmentation," *CoRR*, 2024.

[10] R. Wu, Y. Liu, P. Liang, and Q. Chang, "H-vmunet: High-order vision mamba unet for medical image segmentation," *Neurocomputing*, p. 129447, 2025.

[11] Z. Dai, B. K. H. Low, and P. Jaillet, "Differentially private federated bayesian optimization with distributed exploration," *Advances in Neural Information Processing Systems*, vol. 34, pp. 9125–9139, 2021.

[12] W. Yang, Y. Zhou, M. Hu, D. Wu, X. Zheng, J. H. Wang, S. Guo, and C. Li, "Gain without pain: Offsetting dp-injected noises stealthily in cross-device federated learning," *IEEE Internet of Things Journal*, vol. 9, no. 22, pp. 22 147–22 157, 2021.

[13] J. Chen, J. Mei, X. Li, Y. Lu, Q. Yu, Q. Wei, X. Luo, Y. Xie, E. Adeli, Y. Wang *et al.*, "Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers," *Medical Image Analysis*, vol. 97, p. 103280, 2024.

[14] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-unet: Unet-like pure transformer for medical image segmentation," in *European conference on computer vision*. Springer, 2022, pp. 205–218.

[15] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations*, 2021.

[16] A. Gu and T. Dao, "Mamba: Linear-time sequence modeling with selective state spaces," *Conference on Language Modeling*, 2024.

[17] Z. Wang, J.-Q. Zheng, Y. Zhang, G. Cui, and L. Li, "Mamba-unet: Unet-like pure visual mamba for medical image segmentation," *CoRR*, 2024.

[18] K. Bonawitz, H. Eichner, W. Grieskamp, D. Huba, A. Ingerman, V. Ivanov, C. Kiddon, J. Konečný, S. Mazzocchi, B. McMahan *et al.*, "Towards federated learning at scale: System design," *Proceedings of machine learning and systems*, vol. 1, pp. 374–388, 2019.

[19] L. Gao, H. Fu, L. Li, Y. Chen, M. Xu, and C.-Z. Xu, "Feddc: Federated learning with non-iid data via local drift decoupling and correction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 10 112–10 121.

[20] L. Wang, K. Zhang, Y. Li, Y. Tian, and R. Tedrake, "Does learning from decentralized non-iid unlabeled data benefit from self supervision?" *The Eleventh International Conference on Learning Representations*, 2023.

[21] D. Caldarola, B. Caputo, and M. Ciccone, "Improving generalization in federated learning by seeking flat minima," in *European Conference on Computer Vision*. Springer, 2022, pp. 654–672.

[22] M. Jiang, Z. Wang, and Q. Dou, "Harmofl: Harmonizing local and global drifts in federated learning on heterogeneous medical images," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 1087–1095.

[23] 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.

[24] H.-P. Wang, D. Chen, R. Kerkouche, and M. Fritz, "Fedlap-dp: Federated learning by sharing differentially private loss approximations," in *Proceedings on Privacy Enhancing Technologies*, vol. 2024, no. 3, 2024, pp. 372–390.

[25] P. Tschandl, C. Rosendahl, and H. Kittler, "The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions," *Scientific Data*, vol. 5, p. 180161, 2018. [Online]. Available: https://doi.org/10.1038/sdata.2018.161