SiFT:A Serial Framework with Textual Guidance for Federated Learning

Xuyang Li[1,3], Weizhuo Zhang[1,3], Yue Yu[2], Wei-Shi Zheng[1,3], Tong Zhang[2(✉)], and Ruixuan Wang[1,2,3(✉)]

[1] School of Computer Science and Engineering, Sun Yat-sen Univerisity, Guangzhou, China
[2] Peng Cheng Laboratory, Shenzhen, China
zhangt02@pcl.ac.cn
[3] Key Laboratory of Machine Intelligence and Advanced Computing, MOE, Guangzhou, China
wangruix5@mail.sysu.edu.cn

Abstract. Deep learning has been extensively used in various medical scenarios. However, the data-hungry nature of deep learning poses significant challenges in the medical domain, where data is often private, scarce, and imbalanced. Federated learning emerges as a solution to this paradox. Federated learning aims to collaborate multiple data owners (i.e., clients) for training a unified model without requiring clients to share their private data with others. In this study, we propose an innovative framework called SiFT (**S**er**i**al **F**ramework with **T**extual guidance) for federated learning. In our framework, the model is trained in a cyclic sequential manner inspired by the study of continual learning. In particular, with a continual learning strategy which employs a long-term model and a short-term model to emulate human's long-term and short-term memory, class knowledge across clients can be effectively accumulated through the serial learning process. In addition, one pre-trained biomedical language model is utilized to guide the training of the short-term model by embedding textual prior knowledge of each image class into the classifier head. Experimental evaluations on three public medical image datasets demonstrate that the proposed SiFT achieves superior performance with lower communication cost compared to traditional federated learning methods. The source code is available at https://openi.pcl.ac.cn/OpenMedIA/SiFT.git.

Keywords: Federated learning · Continual learning · Biomedical language model

1 Introduction

Human-level performance of deep learning models is often built on large-scale training data [12, 23]. However, centrally collecting sufficient data remains a challenge due to privacy concerns. This challenge is even more pronounced in the medical domain where medical centers are often reluctant to share their

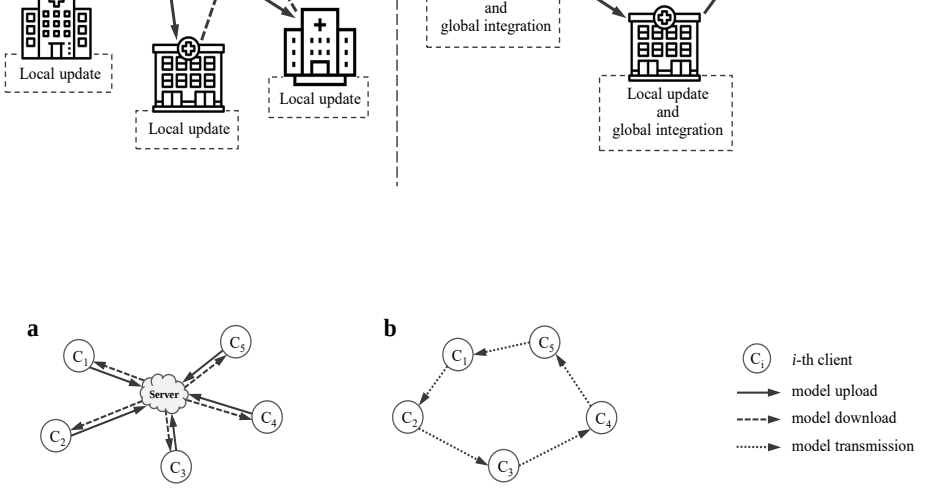


Fig. 1: Comparison between (a) conventional federated learning framework and (b) the proposed serial framework SiFT.

medical data with each other [22, 38]. To solve this challenge, federated learning has been developed by using medical data from multiple medical centers for the training of a unified deep learning model, under the condition that each medical center keeps its own data localized and inaccessible by others [1, 16, 21].

The conventional federated learning framework often involves one server and multiple clients, as illustrated in Figure 1a. Each selected client (e.g., mobile phone or hospital) trains a local model based on its own dataset, and the server aggregates all collected local models to obtain a global model. The global model is then sent back to clients to help refine local models. Such a process is repeated multiple times to achieve a final global model which is expected to perform comparably well as a model trained on the collection of all clients' data. Representative federated learning methods include FedAvg [21], SCAFFOLD [11], FedDyn [1], FedSpeed [27], etc. While these methods work well in the simple independent and identically distributed (IID) setting for data distributions across clients, their performance often degrades with slower convergence when data distributions vary substantially across clients due to client drift [11, 28]. Researchers attempt to mitigate the effect of client drift from multiple aspects, e.g., by designing proximal terms [1, 16, 27], improving local model generalization [15, 40], proposing novel aggregation strategies [8, 32], or constructing a personalized model for each client [24, 29], etc. In addition, particularly in the medical imaging domain, researchers started to explore federated learning under self-supervised or semi-supervised settings [18, 25, 34]. Decentralized federated learning(DFL) eliminates the need for a central server [4, 19]. But most methods for DFL still need model synchronization, which leads to high communication overhead and vulnerability to client drift. Some DFL methods adopt ring-like network topology [19, 33], but are still limited by the model synchronization. A few methods that involve cyclically serial training have been proposed [3, 6], but they are only effective in (near) IID scenarios.

In this study, we propose a new federated learning framework SiFT to learn a unified image classification model in a serial manner (Figure 1b). Different from the general federated learning framework which requires a server to communicate with multiple clients in a parallel manner, the proposed serial framework does not require the server and sequentially updates the model across clients. Such a serial federated learning framework is inspired by the study of continual learning which aims to sequentially learn new knowledge and simultaneously preserve old

knowledge [17, 36, 39]. In this study, a continual learning strategy based on the complementary learning system (CLS) theory [13] is utilized. Specifically, two complementary models are employed, referred to as long-term model and short-term model respectively. The long-term model is considered as the long-term memory and short-term model is as the short-term memory of CLS. In particular, the short-term model (i.e., image classifier) is trained under the guidance of one pre-trained biomedical language model, such that textual prior knowledge of medical image classes can be effectively utilized as a train-free classifier head to help train the feature extractor of the short-term model. Empirical evaluations on multiple medical image datasets confirm the superior performance of the proposed SiFT framework compared to representative federated learning methods. Our contributions are summarized below.

- A new and effective serial framework for federated learning that elegantly integrates the CLS theory and is guided by a language model;
- A continual learning strategy based on the CLS theory to sequentially accumulate knowledge across clients for federated learning;
- Innovatively utilizing prior knowledge from a biomedical language model as a train-free classifier head to help train the image classifier.

2 Methodology

The objective of federated learning is to train a global model with datasets of multiple clients, under the condition that the dataset of each client is private and therefore kept localized. Different from the traditional parallel framework which uses a server to aggregate parameters of local models from multiple clients, a novel and effective serial framework for federated learning is proposed here without requiring the server.

2.1 Overview of the Serial Framework

The proposed federated learning framework works in a serial manner (Figure 1b), i.e., each client receives the model from the preceding client, updates the model with a local dataset, and then sends the updated model to the next client. Such a serial process can be cyclically performed, e.g., the updated model from the last client can be sent to the first client to start a new round of model training. In this serial framework, a continual learning strategy inspired by the complementary learning system (CLS) theory [13] is innovatively applied to help the model effectively accumulate knowledge across clients. The CLS theory [13] claims that a short-term memory is used to learn new knowledge and then a long-term memory is updated to assimilate new knowledge temporarily stored in the short-term memory. Accordingly, a long-term model and a short-term model are utilized in the proposed continual learning strategy. The long-term model needs no training and is updated by the short-term model (Figure 2a). Such a serial learning framework is expected to be more efficient and effective than the

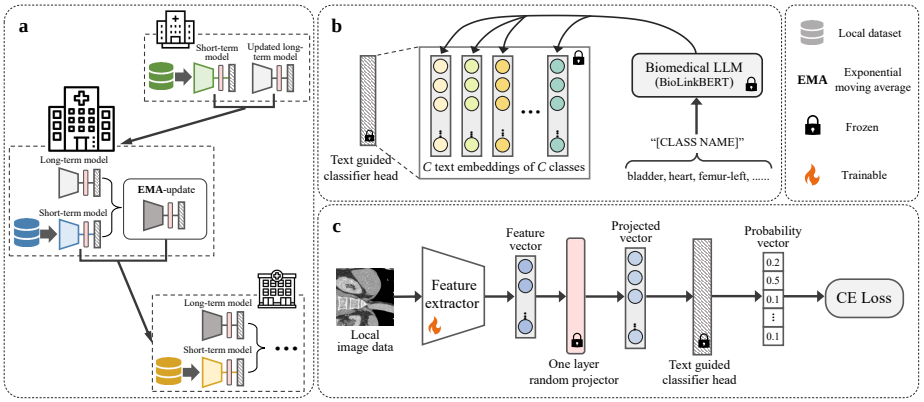


Fig. 2: The proposed serial framework for federated learning. (a) Each client updates the short-term model using its local image data, followed by updating the long-term model through the EMA strategy, and serially sending the updated model pair to the next client. (b) Text embeddings of all classes from the pre-trained biomedical language model form the text-guided train-free classifier head. (c) The feature extractor of the short-term model is trained under the guidance of the text-guided classifier head.

aggregation of all collected local models in the conventional parallel federated learning paradigm, particularly when the data distributions vary significantly across diverse clients. In particular, the short-term model is trained under the guidance of textual prior knowledge of all classes. The training of the short-term model and the update of the long-term model are detailed below.

2.2 Short-term Model with Text-guided Train-free Classifier Head

The short-term model received from the preceding client will be updated with the dataset of the current client. For the m-th client, the short-term model (i.e., image classifier) includes a trainable feature extractor $F_m(\cdot)$, a frozen projector $P(\cdot)$, and a text-guided train-free classifier head $H(\cdot)$. While the feature extractor F_m conventionally consists of multiple convolutional layers, the projector P and the classifier head H are innovatively designed.

Specially, a pre-trained and fixed biomedical language model is utilized to construct the classifier head H(Figure 2b). Inspired by the successful applications of vision-language model CLIP [23] to natural image classification, we hypothesize that prior knowledge of medical categories in the pre-trained biomedical language model may help the medical image classifier more effectively learn the knowledge from medical images. Here, the prior knowledge of categories and the categorical relationships learned by the biomedical language models are naturally embedded into the classifier head of the image classifier as follows. Suppose there are totally C medical image classes to be learned by the image classifier, and denote by $\mathbf{t}_c$ the textual description of the c-th image class. $\mathbf{t}_c$ can

be simply the class name or a rich description of the class by querying a large language model like GPT-4 [2] or consulting a medical expert. With $\mathbf{t}_c$ as the input to the pre-trained biomedical language model, the corresponding output vector $\mathbf{w}_c \in \mathbb{R}^d$ from the biomedical language model is collected and utilized as part of the weight parameters in the classifier head. After collecting all the class embeddings, the weight matrix $\mathbf{W}$ of the classifier head is obtained as $\mathbf{W} = [\mathbf{w}_1\, \mathbf{w}_2\, \ldots\, \mathbf{w}_C] \in \mathbb{R}^{d\times C}$, where d is the length of the input to the classifier head and also the dimension of the output vector from the biomedical language model. Note that the weight matrix $\mathbf{W}$ contains the textual prior knowledge of the C image classes obtained from the pre-trained biomedical language model. In order to preserve the textual prior knowledge from being altered, the constructed classifier head remains frozen and undergoes no changes during the training process of the short-term model.

Considering that the input dimension d of the classifier head is pre-determined by the output dimension of the biomedical language model and therefore may be different from the output dimension D of the image feature extractor F, the projector P is introduced to help align image representations with the textual representations. Inspired by successful applications of random projection [20, 26], here a randomly initialized and frozen matrix with shape $D \times d$ is used to represent the projector, which is proved to be better than a trainable projector (see Figure 4d in the ablation study).

Then, given any input image $\mathbf{x}$ to the image classifier, the output of the classifier can be obtained by

$$p_c(\mathbf{x}) = \frac{\exp\left(\langle \mathbf{w}_c, P(F_m(\mathbf{x}))\rangle / \tau\right)}{\sum_{k=1}^{C} \exp\left(\langle \mathbf{w}_k, P(F_m(\mathbf{x}))\rangle / \tau\right)}, \tag{1}$$

where $p_c(\mathbf{x})$ denotes the c-th element of the image classifier output $\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x})\, p_1(\mathbf{x})\, \ldots\, p_C(\mathbf{x})]^\intercal$, and $\langle \mathbf{a}, \mathbf{b}\rangle$ represents the cosine similarity between any two vectors $\mathbf{a}$ and $\mathbf{b}$. The cosine similarity is used here following the applications of vision-language models [23]. τ is the temperature coefficient. With the help of the frozen projector P and the train-free classifier head H, the trainable feature extractor F_m can be easily trained by the cross-entropy loss over the training set of each client (Figure 2c).

2.3 Update of the Long-term Model

The long-term model is also sent from the preceding $(m-1)$-th client to the current m-th client for update. Based on the CLS theory [13], the trained short-term model representing the short-term memory at the m-th client will be used to update the long-term model (i.e., long-term memory). Here suppose the long-term model shares the same frozen projector P and classifier head H with short-term model. Therefore, only the feature extractor of the long-term model needs to be updated. Also suppose the feature extractor of the long-term model has the same structure as that of the short-term model. As in general continual learning [5, 31], the aim of updating the long-term model is to learn new knowledge and

preserve old knowledge simultaneously. Following the strategy of accumulating old knowledge in self-supervised learning methods like MoCo [9], the exponential moving average (EMA) is utilized to update the long-term model. Formally, denote by $\boldsymbol{\theta}_m$ the parameters of the trained short-term model at the m-th client, and $\boldsymbol{\theta}$ the parameters of the long-term model. Then the long-term model can be updated by EMA as follows,

$$\boldsymbol{\theta} \leftarrow \beta\boldsymbol{\theta} + (1-\beta)\boldsymbol{\theta}_m\,, \tag{2}$$

where the hyperparameter $\beta \in (0, 1)$ is the smoothing factor of EMA and often set a higher value (e.g., 0.9) to better preserve old knowledge in the long-term model.

During inference, the long-term model (i.e., feature extractor updated from the short-term model followed by the frozen projector and the text-guided train-free classifier head) can be used to predict the class of any test image.

2.4 Experimental Setup

Datasets. Our method was evaluated on three widely-used medical datasets, including HAM10000 [30], OrganCMNIST [35], and OrganSMNIST [35] (See Table 1 in the Supplementary Material for dataset details). Note that HAM10000 is a highly class-imbalanced dataset.

Data distributions. By default, suppose there are totally 10 clients, and each training dataset is partitioned into 10 subsets, with one subset for one client. For the HAM10000 [30], given its highly class-imbalanced distribution property, each client's data distribution adheres to the independent and identically distributed (IID) setting. With OrganCMNIST [35] and OrganSMNIST [35], to more realistically simulate real-world data settings following previous studies [14], the Dirichlet distribution is employed to adjust the distribution of sample size over clients for each class. Specifically, we set the Dirichlet distribution coefficient α to 0.5 and 0.3 respectively for each of the two datasets. A smaller coefficient α indicates stronger level of data heterogeneity across the clients for each class.

Implementation. Following previous studies [1, 7, 27] and for fair comparisons, ResNet18 [10] was employed as the model backbone in all experiments. We adapted the BioLinkBERT-large [37] as the biomedical language model to construct the train-free classifier head, where the textual class names were merely used as semantic information by default. For all federated learning methods, one federated learning round refers to the completion of a learning iteration where all clients have participated, and local learning epochs refer to the epochs to learn the model at each client within a learning round. SGD optimizer was used for model training, with initial learning rate 0.1 for the first learning round and reduced to 0.01 for subsequent rounds. For existing federated learning methods, the initial learning rate was set to 0.1 and then decayed with factor 0.998 over learning rounds. For all methods, batch size was 128, and unless mentioned otherwise, local learning epochs were set to 10. Given the divergent convergence rates among different methods, the number of learning rounds was set to 300 for

Table 1: Performance comparison in balanced accuracy (%). "Dir α" denotes the training data distribution follows Dirichlet distribution and the coefficient is α. Mean and standard deviation over three runs are reported.

Method	HAM10000	OrganCMNIST		OrganSMNIST	
	IID	Dir 0.5	Dir 0.3	Dir 0.5	Dir 0.3
FedAvg[21]	56.602±1.90	89.288±0.44	88.992±0.38	71.338±0.07	69.381±0.91
FedProx[16]	58.456±1.50	89.296±0.29	88.820±0.23	71.527±0.32	70.213±0.36
SCAFFOLD[11]	58.018±1.41	88.986±0.53	88.242±0.04	69.685±0.53	67.307±1.08
FedDyn[1]	57.076±2.91	89.435±0.31	88.594±0.16	72.285±0.36	72.319±1.10
FedSpeed[27]	55.723±1.58	89.458±0.87	90.049±0.44	72.784±0.92	73.927±0.16
SiFT	**63.625±1.78**	**91.522±0.10**	**90.234±0.25**	**74.825±0.65**	**74.595±0.48**

Table 2: The number of training rounds required to achieve the target balanced accuracy.

Method	HAM10000 IID				OrganCMINST Dir0.5				OrganCMINST Dir0.3			
	25%	35%	45%	55%	55%	65%	75%	85%	55%	65%	75%	85%
FedAvg[21]	22	42	50	158	24	34	48	104	31	39	60	123
FedProx[16]	29	44	63	158	25	33	52	92	28	39	59	113
SCAFFOLD[11]	29	44	68	174	28	41	59	144	38	51	70	148
FedDyn[1]	19	38	45	97	13	16	24	44	15	21	30	78
FedSpeed[27]	63	114	145	275	30	39	66	119	34	44	64	116
SiFT	**3**	**5**	**7**	**12**	**2**	**3**	**5**	**12**	**8**	**13**	**20**	**23**

existing methods to ensure convergence, while only set to 50 for the proposed SiFT due to its faster convergence. Balanced accuracy, i.e., average of recall over all classes, was adapted as the evaluation metric due to its ability to assess model performance in the presence of imbalanced data distributions across classes.

2.5 Experiment Results

Effectiveness evaluation. To demonstrate the effectiveness of the proposed SiFT, a variety of traditional federated learning methods were used for comparison, including FedAvg [21], FedProx [16], FedDyn [1], SCAFFOLD [11] and FedSpeed [27]. As shown in Table 1, SiFT outperforms all the federated learning baselines on all the datasets with various settings. The performance improvement is more significant on HAM10000 probably because the dataset is highly class-imbalanced compared to the others. These results support that the proposed SiFT framework is a strong alternative to existing client-server based framework for federated learning.

Convergence speed. In addition to classification performance, convergence speed is also concerned in federated learning, as faster convergence often implies lower communication cost. As shown in Table 2, to achieve the same level of classification performance, SiFT requires much fewer training rounds, e.g., around

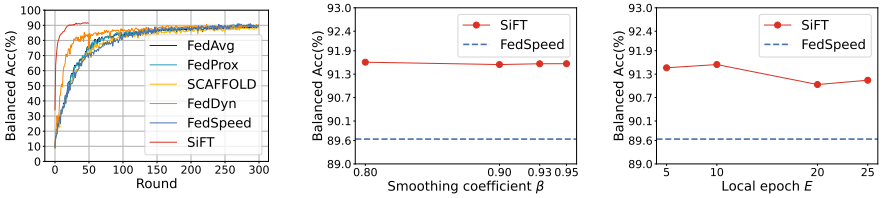
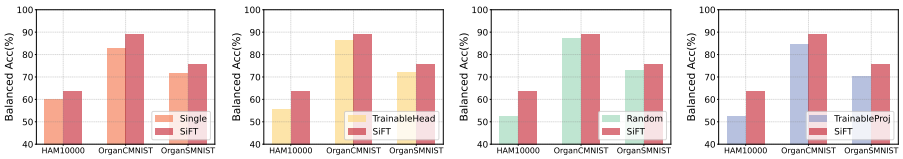


Fig. 3: Performance of various methods on the test set of OrganCMNIST over learning rounds (left, $\alpha = 0.5$) and sensitivity study of two hyper-parameters (middle and right).



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

Fig. 4: Ablation study of the proposed method on HAM10000 (IID), OrganCMNIST ($\alpha = 0.3$) and OrganSMNIST ($\alpha = 0.5$).

only one-seventh of the rounds required to achieve similar performance compared to the highly performing FedDyn [1] on HAM10000. This suggests that SiFT converges significantly faster than the top-performing federated learning methods, which also implies a substantial reduction in communication cost. This is also confirmed in Figure 3 (left) which demonstrates the evolution of the balanced accuracy on one test set for each method during the training process.

Sensitivity study. The proposed SiFT is largely insensitive to the value choice of the smoothing coefficient β in EMA (Equation 2), as demonstrated in Figure 3 (middle). In addition, as shown in Figure 3 (right), when varying the number of local training epochs E (but keeping the multiplication of E and federated learning rounds constant), the final long-term model performs stably well.

Ablation study. Ablation studies were performed to confirm the effect of each key component in the proposed SiFT framework. When removing the long-term model ('Single'), which means only short-term model is updated sequentially and evaluated, the balanced accuracy degrades clearly (Figure 4a), supporting the necessity of using two complementary networks. When replacing the frozen projector and the train-free classifier head by a trainable classifier head without the guidance of the biomedical language model ('TrainableHead'), or replacing the train-free classifier head by a randomly initialized and fixed classifier head ('Random'), the final long-term model again performs worse (Figure 4b&c), confirming the effect of the text-guided train-free classifier head. In addition, replacing the frozen projector by a trainable projector ('TrainableProj') also results in

degraded performance (Figure 4d), probably because trainable projector introduces more model parameters which more likely cause over-fitting.

3 Conclusion

This paper proposes a novel serial framework called SiFT for federated learning. In this framework, the paired short-term and long-term models are sequentially updated across clients to accumulate knowledge. The short-term model is effectively trained and updated with the guidance of textual prior knowledge from a pre-trained biomedical language model, and the long-term model is updated by the short-term model based on the exponential moving average strategy at each client. Extensive evaluations on three medical image datasets confirm the effectiveness and communication efficiency of the proposed framework. More effective continual learning strategies and textual prior knowledge from richer textual description may further improve the efficacy of the serial federated learning framework, which will be investigated as future work.

Acknowledgments. This work is supported in part by the National Natural Science Foundation of China (grant No. 62071502), the Major Key Project of PCL (grant No. PCL2023A09), and Guangdong Excellent Youth Team Program (grant No. 2023B1515040025).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Acar, D.A.E., Zhao, Y., Matas, R., Mattina, M., Whatmough, P., Saligrama, V.: Federated learning based on dynamic regularization. In: ICLR (2020)
2. Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F.L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., et al.: Gpt-4 technical report. arXiv preprint arXiv:2303.08774 (2023)
3. Balachandar, N., Chang, K., Kalpathy-Cramer, J., Rubin, D.L.: Accounting for data variability in multi-institutional distributed deep learning for medical imaging. JAMIA (2020)
4. Beltrán, E.T.M., Pérez, M.Q., Sánchez, P.M.S., Bernal, S.L., Bovet, G., Pérez, M.G., Pérez, G.M., Celdrán, A.H.: Decentralized federated learning: Fundamentals, state of the art, frameworks, trends, and challenges. IEEE Communications Surveys & Tutorials (2023)
5. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: ECCV (2018)
6. Chang, K., Balachandar, N., Lam, C., Yi, D., Brown, J., Beers, A., Rosen, B., Rubin, D.L., Kalpathy-Cramer, J.: Distributed deep learning networks among institutions for medical imaging. JAMIA (2018)

7. Chen, M., Jiang, M., Dou, Q., Wang, Z., Li, X.: Fedsoup: Improving generalization and personalization in federated learning via selective model interpolation. In: MICCAI (2023)
8. Deng, Z., Li, D., Tan, S., Fu, Y., Yuan, X., Huang, X., Zhang, Y., Zhou, G.: Fedgrav: An adaptive federated aggregation algorithm for multi-institutional medical image segmentation. In: MICCAI (2023)
9. He, K., Fan, H., Wu, Y., Xie, S., Girshick, R.: Momentum contrast for unsupervised visual representation learning. In: CVPR (2020)
10. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: CVPR (2016)
11. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S., Stich, S., Suresh, A.T.: Scaffold: Stochastic controlled averaging for federated learning. In: ICML (2020)
12. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)
13. Kumaran, D., Hassabis, D., McClelland, J.L.: What learning systems do intelligent agents need? complementary learning systems theory updated. Trends in Cognitive Sciences (2016)
14. Li, Q., Diao, Y., Chen, Q., He, B.: Federated learning on non-iid data silos: An experimental study. In: ICDE (2022)
15. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: CVPR (2021)
16. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. MLSys (2020)
17. Li, Z., Hoiem, D.: Learning without forgetting. TPAMI (2017)
18. Liu, Q., Yang, H., Dou, Q., Heng, P.A.: Federated semi-supervised medical image classification via inter-client relation matching. In: MICCAI (2021)
19. Liu, W., Chen, L., Zhang, W.: Decentralized federated learning: Balancing communication and computing costs. IEEE T-SIPN (2022)
20. McDonnell, M., Gong, D., Parvaneh, A., Abbasnejad, E., van den Hengel, A.: Ranpac: Random projections and pre-trained models for continual learning. In: NeurIPS (2023)
21. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: AISTATS (2017)
22. Pfitzner, B., Steckhan, N., Arnrich, B.: Federated learning in a medical context: a systematic literature review. TOIT (2021)
23. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: ICML (2021)
24. Roth, H.R., Yang, D., Li, W., Myronenko, A., Zhu, W., Xu, Z., Wang, X., Xu, D.: Federated whole prostate segmentation in mri with personalized neural architectures. In: MICCAI (2021)
25. Saha, P., Mishra, D., Noble, J.A.: Rethinking semi-supervised federated learning: How to co-train fully-labeled and fully-unlabeled client imaging data. In: MICCAI (2023)
26. Schmidt, W.F., Kraaijveld, M.A., Duin, R.P., et al.: Feed forward neural networks with random weights. In: ICPR (1992)
27. Sun, Y., Shen, L., Huang, T., Ding, L., Tao, D.: Fedspeed: Larger local interval, less communication round, and higher generalization accuracy. In: ICLR (2022)
28. Tan, A.Z., Yu, H., Cui, L., Yang, Q.: Towards personalized federated learning. TNNLS (2022)

29. Tarumi, S., Suzuki, M., Yoshida, H., Miyauchi, S., Kurazume, R.: Personalized federated learning for institutional prediction model using electronic health records: A covariate adjustment approach. In: EMBC (2023)
30. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data (2018)
31. Wang, F.Y., Zhou, D.W., Ye, H.J., Zhan, D.C.: Foster: Feature boosting and compression for class-incremental learning. In: ECCV (2022)
32. Wang, M., Wang, L., Xu, X., Zou, K., Qian, Y., Goh, R.S.M., Liu, Y., Fu, H.: Federated uncertainty-aware aggregation for fundus diabetic retinopathy staging. MICCAI (2023)
33. Wang, Z., Hu, Y., Yan, S., Wang, Z., Hou, R., Wu, C.: Efficient ring-topology decentralized federated learning with deep generative models for medical data in ehealthcare systems. Electronics (2022)
34. Wu, Y., Zeng, D., Wang, Z., Shi, Y., Hu, J.: Federated contrastive learning for volumetric medical image segmentation. In: MICCAI (2021)
35. Yang, J., Shi, R., Wei, D., Liu, Z., Zhao, L., Ke, B., Pfister, H., Ni, B.: Medmnist v2-a large-scale lightweight benchmark for 2d and 3d biomedical image classification. Scientific Data (2023)
36. Yang, Y., Cui, Z., Xu, J., Zhong, C., Zheng, W.S., Wang, R.: Continual learning with bayesian model based on a fixed pre-trained feature extractor. Visual Intelligence (2023)
37. Yasunaga, M., Leskovec, J., Liang, P.: Linkbert: Pretraining language models with document links. arXiv preprint arXiv:2203.15827 (2022)
38. Yuan, L., Liu, X., Yu, J., Li, Y.: A full-set tooth segmentation model based on improved pointnet++. Visual Intelligence (2023)
39. Zhang, W., Huang, Y., Zhang, T., Zou, Q., Zheng, W.S., Wang, R.: Adapter learning in pretrained feature extractor for continual learning of diseases. In: MICCAI (2023)
40. Zhou, Q., Zheng, G.: Fedcontrast-gpa: Heterogeneous federated optimization via local contrastive learning and global process-aware aggregation. In: MICCAI (2023)