Bridging Knowledge Discrepancy in Retinal Image Analysis through Federated Multi-Task Learning

Jing Yang[1], Yuxi Ma[1], Jin-Gang Yu[3], Feng Gao[4], Shuting Yang[4], Du Cai[4], Jiacheng Wang[5(✉)], and Liansheng Wang[1,2(✉)]

[1] National Institute for Data Science in Health and Medicine, Xiamen University, Xiamen 361005, China
jingy77@stu.xmu.edu.cn, mayuxi1@stu.xmu.edu.cn
[2] Department of Computer Science at the School of Informatics, Xiamen University, Xiamen 361005, China
lswang@xmu.edu.cn
[3] School of Automation Science and Engineering, South China University of Technology, Guangzhou 510641, China
jingangyu@scut.edu.cn
[4] The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou 510655, China
gaof57@mail.sysu.edu.cn, yangsht35@mail.sysu.edu.cn, caid28@mail.sysu.edu.cn
[5] Manteia Technology Co., Ltd, Xiamen 361005, China
jiachengw@stu.xmu.edu.cn

Abstract. Retinal image analysis not only reveals the microscopic structure of the eye but also provides insights into overall health status. Therefore, employing multi-task learning to simultaneously address disease recognition and segmentation in retinal images can improve the accuracy and comprehensiveness of the analysis. Given the need for medical privacy, federated multi-task learning provides an effective solution for retinal image analysis. However, existing federated multi-task learning studies fail to address client resource constraints or knowledge discrepancies between global and local models. To address these challenges, we propose FedBKD, a novel federated multi-task learning framework for retinal image analysis. FedBKD leverages a server-side foundation model and effectively bridges the knowledge discrepancy between the clients and the server. Before local training, the adaptive sub-model extraction module ranks the activation values of neurons in the global model. It extracts the most representative sub-model based on computational resources, thereby facilitating the local adaptation of the global model. Additionally, we design a feature consistency optimization strategy to ensure alignment between the local model and the global foundation model's prior knowledge. This reduces error accumulation in the client sub-model during multi-task learning and ensures better adaptation to local tasks. Experimental results on the multi-center retinal image dataset demonstrate that FedBKD achieves state-of-the-art performance. Our code is available at https://github.com/Yjing07/FedBKD.git.

Keywords: Federated Multi-Task Learning · Retinal Image Analysis · Knowledge Discrepancy.

1 Introduction

With the rapid development of artificial intelligence, foundation models (FM) have shown great potential in the medical field [19, 11], especially in retinal image analysis. For instance, RETFound [21], trained on over 1.6 million retinal images, has demonstrated excellent performance in diagnostic and prognostic tasks of ocular diseases, highlighting its powerful image understanding capabilities. Retinal images provide detailed information about the microstructure of the eye but also indicate diseases linked to overall health. Symptoms and lesions of these diseases can manifest in different regions and layers, requiring multi-task learning for accurate recognition, segmentation, and prediction. However, individual medical institutions face challenges like lacking annotated data and resource limitations. At the same time, considering privacy protection and data security, developing a solution that enables collaborative training across multiple medical centers while ensuring data privacy and enhancing the application capabilities of FM has become an important research direction with vast potential.

Federated Multi-Task Learning (FMTL) combines the advantages of federated learning (FL) and multi-task learning by centrally aggregating model updates, reducing the risk of sensitive data leakage. It enables the model to learn multiple related tasks simultaneously, sharing common features and significantly improving task performance. For instance, MaT-FL [3] and FedHCA^2 [15] optimize task collaboration efficiency by dynamically adjusting client grouping or strategies. However, these approaches do not adequately address resource limits among different clients. Studies like FedDrop [4, 6], HeteroFL [7, 10], and FedRolex [1] have introduced strategies that randomly, statically, or rolling select sub-models of the global model for local training. However, they also present two key challenges. Firstly, discarding specific neurons to reduce computational and communication burdens does not fully adopt the global model. Directly removing neurons from the FM may cause the sub-model to lose some task-relevant features, particularly in segmentation tasks. Secondly, multi-task learning for retinal images requires the model to simultaneously handle functions with different features. Due to the knowledge discrepancy between the sub-model and the global model, gradient errors accumulate during training, affecting the model's accuracy and stability.

To address these challenges, we propose a novel federated multi-task learning framework for retinal image analysis with **B**ridging **K**nowledge **D**iscrepancy (FedBKD), which integrates two key components: the Adaptive Sub-model Extraction (ASE) module and the Feature Consistency-based Optimization (FCO) strategy. The ASE module ranks the activation values of neurons in the server-side FM. It extracts the most representative sub-model for each client, facilitating the local adaptation of the global model. This module ensures that the sub-model focuses on the most significant features of local tasks, thereby enhanc-

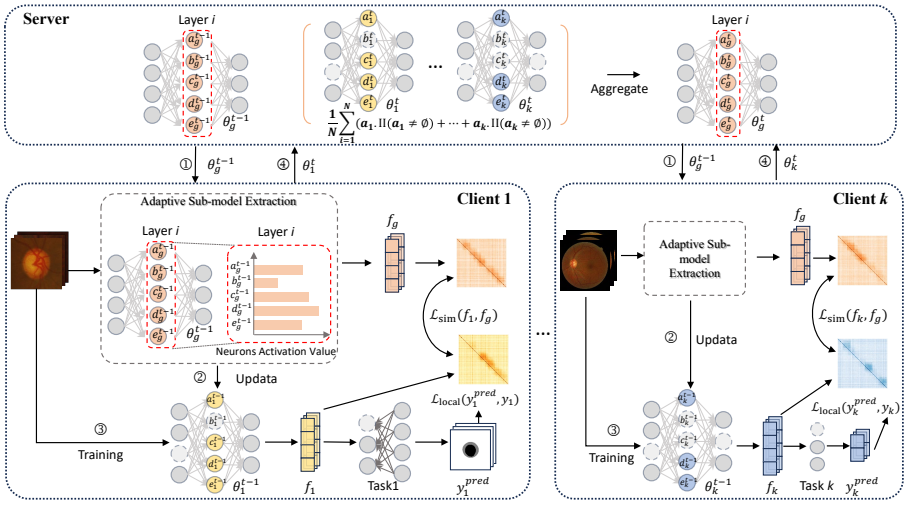

Fig. 1. The workflow of FedBKD. Each client first localizes the received encoder using ASE (Section 2.2). Then, the client performs local training and global parameter updates using the FCO (Section 2.3).

ing the adaptability of the FM across different tasks. The FCO strategy optimizes feature consistency between the global and local models, ensuring feature consistency from client training and global parameter aggregation perspectives. This strategy effectively bridges the knowledge discrepancy caused by feature differences in retinal images, further reducing gradient errors and improving the overall performance and stability of the sub-model.

Our main contributions are summarized as follows: (1) We introduce FedBKD, a novel FMTL method that uses FM to assist in training different computing clients and dynamically optimize knowledge discrepancy. FedBKD enhances model adaptability in heterogeneous client environments by adaptively extracting key neurons from the global model. (2) We introduce a feature consistency optimization strategy that aligns the client sub-model with the global model, ensuring effective knowledge transfer and significantly improving task performance. (3) We validate FedBKD on the multi-center retinal dataset and achieve state-of-the-art (SOTA) performance in multi-task scenarios.

2 Methods

2.1 Preliminary

Given K clients, each client k with a local dataset $D_k = \{(x_k^i, y_k^i), i = 1, 2, \ldots, N_k\}$, where N_k is the number of samples, the tasks are private to each client. The local model parameters for client k are denoted as θ_k, and its model capacity β_k represents the proportion of neurons extracted from the global model, where

$0 < \beta_k \leq 1$. During training, each client performs local updates using private data. Afterward, the server aggregates the encoder parameters from all clients to improve the global model, while task-specific decoders are excluded from the global update. The overall framework of our FedBKD is shown in Fig. 1, and the pipeline is detailed in Algorithm 1.

2.2 Adaptive Sub-Model Extraction

Due to the limitations in computation and storage on the client side, it cannot support the training tasks of a large-scale global model. In contrast, the server side has more substantial computational resources, making it a key challenge to extract a small model suitable for the client from the large model. This study proposes an adaptive sub-model selection method based on neuron activation values, which extracts a task-specific sub-model from the FM while maximizing the utilization of client resources.

Each Transformer layer comprises two key components: Multi-Head Self-Attention (MHA) and Feed-Forward Network (FFN). To filter the most influential neurons, we calculate the activation values and importance of the global model's neurons, considering the client's model capacity. The output of a specific layer is then obtained through the FFN, defined as:

$$\theta_{\mathrm{FFN}}(x) = W_2 \cdot \mathrm{gelu}(xW_1 + b_1) + b_2, \tag{1}$$

where $W_1 \in \mathbb{R}^{d_{\mathrm{FFN}} \times d_{\mathrm{model}}}$ and $W_2 \in \mathbb{R}^{d_{\mathrm{model}} \times d_{\mathrm{FFN}}}$ are weight matrices, and b_1 and b_2 are bias, d_{FFN} is the dimensions of FFN hidden layer and set $4 \times d_{\mathrm{model}}$. Here, the activation function $\mathrm{gelu}(\cdot)$ is applied element-wise to the result of the linear transformation $xW_1 + b_1$. To further assess the importance of the neurons, we combine the mean and variance of the activations, proposing the following saliency measure:

$$\psi(m) = (\alpha(m))^2 + \gamma(m), \tag{2}$$

where:

$$\begin{cases} \alpha(m) = \frac{1}{N_k} \sum_{j=1}^{N_k} \left(\mathrm{gelu}(xW_1 + b_1)_m\right), \\ \gamma(m) = \frac{1}{N_k} \sum_{j=1}^{N_k} \left(\mathrm{gelu}(xW_1 + b_1)_m - \alpha(m)\right)^2. \end{cases} \tag{3}$$

Here, $\alpha(m)$ represents the mean activation value of the m-th neuron across all samples, and $\gamma(m)$ represents the variance of the activation values for the m-th neuron. The top β_k neurons are selected based on the saliency measure $\psi(m)$, and a sub-model is constructed using these neurons. By adjusting the client's parameters, we ensure that sub-model is tailored to the specific task requirements of each client while effectively reducing the computational burden on client devices.

2.3 Feature Consistency-based Optimization Strategy

Due to the significant capacity difference between server and client models, feature inconsistencies lead to gradient error accumulation during the training process. We propose a FCO strategy to tackle this challenge from both local training

Algorithm 1: FedBKD

Input: Local datasets $\{D_k\}_{k=1}^{K}$, Server parameters θ_s, total communication rounds T, learning rate η.

Output: Trained models $\theta^{(T)} = \left\{\theta_1^{(T)}, \ldots, \theta_K^{(T)}\right\}$, $\theta_s^{(T)}$.

for $t = 1$ ***to*** T **do**
  Clients initialize models $\theta^{(t)} = \left\{\theta_1^{(t)}, \ldots, \theta_K^{(t)}\right\}$ with Eq. 2
  for $k = 1$ ***to*** K **do**
    Compute losses L_{total} with Eq. 5
    $(\theta_k^{T+1}, \omega_k^{t+1}) \leftarrow (\theta_k^t, \omega_k^t) - \eta \cdot \nabla L_{\text{total}}(W_K^t)$
    Client uploads its parameters: θ_k^{t+1}
  Server parameter status update: $\frac{1}{\sum w_k} \sum_{k=1}^{K} \theta_k^{t+1} \text{II}(\theta_k^{t+1})$;

and global aggregation perspectives. Local training feature consistency aims to bridge the knowledge discrepancy between the global and local models. Global aggregation feature consistency extracts consensus from task-specific clients to mitigate task discrepancies.

Local training feature consistency: Center Kernel Alignment (CKA) is widely used to measure representation similarity [22]. Inspired by this, we optimize representation similarity by leveraging the guidance of the global model, which helps reduce knowledge discrepancies, mitigates the accumulation of gradient errors, and enhances the accuracy of the client model.

We define f_k as the feature matrix for the k-th client and the f_g for the global feature. The Gram matrices $f_k^T f_k$ and $f_g^T f_g$ capture the intrinsic structure and similarity between the models, and the similarity matrix between these matrices is given by:

$$\mathcal{L}_{\text{sim}}\left(f_k, f_g\right) = \frac{\left\|f_g^T f_k\right\|_F^2}{\left\|f_k^T f_k\right\|_F \left\|f_g^T f_g\right\|_F}, \tag{4}$$

where $\|\cdot\|_F$ represents the Frobenius norm [9], which ensures that the comparison of matrices is unaffected by scale differences, focusing on their structural similarities. We incorporate this feature alignment loss into the original loss function, resulting in the total objective. The final objective function is:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{local}}\left(f_k, y\right) + \lambda \cdot \mathcal{L}_{\text{sim}}\left(f_k, f_g\right), \tag{5}$$

where $\mathcal{L}_{\text{local}}$ is the local loss, and λ is a hyperparameter controlling the weight of the feature alignment.

Global aggregation feature consistency: Inspired by LG-Mix [12], we propose an eigenvalues weighted aggregation strategy to mitigate feature heterogeneity's impact on global model consistency. We first center each client's feature matrix to remove bias, then compute the covariance matrix and apply singular value decomposition to obtain eigenvalues. These eigenvalues are used to weight the features, giving different importance to each sub-model during aggregation. It reduces feature heterogeneity's negative impact and improves the

global model's consistency and robustness, especially in resource-constrained scenarios. Formally, the covariance matrix $\mathbf{C}_k$ is computed and eigenvalue decomposition yields eigenvalues μ_k and eigenvectors $\mathbf{v}_k$:

$$\begin{cases} \mathbf{C}_k = \frac{1}{N_k}\left(\mathbf{X}_k - \frac{1}{N_k}\sum_{k=1}^{N_k}\mathbf{x}_{k,n}\right)\left(\mathbf{X}_k - \frac{1}{N_k}\sum_{k=1}^{N_k}\mathbf{x}_{k,n}\right)^T, \\ \mathbf{C}_k\mathbf{v}_k = \mu_k\mathbf{v}_k, \end{cases} \tag{6}$$

where $\mathbf{x}_{k,n}$ is feature of the k-th client. We then assign weights w_k based on the normalized eigenvalues:

$$w_k = \frac{\mu_k}{\sum_{k=1}^{K}\mu_k}, \tag{7}$$

This weight represents the client's relative importance in the aggregation process. A higher weight for a client indicates a more significant contribution to the global model.

Table 1. Comparison to representative methods using multi-center retinal datasets for five clients. Δ_m represents the average change in performance for each task compared to local training (The client only uses local data for local training without FL).

Method	Segmention					Classification	
	$\beta = 1/16$	$\beta = 1/8$	$\beta = 1/4$	$\beta = 1/2$	average	$\beta = 1/2$	average
	Dice	Dice	Dice	Dice	Δ_m	Acc	Δ_m
Local	79.16	79.40	88.95	90.53	0.0	71.71	0.0
FedProx [14]	82.98	79.02	89.54	91.44	+1.23	72.99	+1.28
Ditto [13]	82.13	80.51	89.73	91.26	+1.40	73.52	+1.81
FedDrop [4]	75.37	74.78	87.82	89.68	-2.60	72.06	+0.35
FedRolex [1]	76.08	75.32	89.18	90.44	-1.76	72.26	-6.45
HeteroFL [7]	82.15	80.65	89.65	90.87	+1.32	73.90	+2.19
MaT-FL [3]	80.61	81.15	89.38	89.15	+0.56	73.71	+2.00
FedHCA^2 [15]	81.02	81.72	89.85	90.46	+1.25	72.24	+1.63
FedBKD (Ours)	**83.54**	**84.14**	**90.06**	**92.21**	+2.98	**75.92**	+4.21

3 Experiments

3.1 Datasets and Implementation

Datasets We conduct experiments on public datasets collected from five different centers. For the segmentation task [2, 16, 17], we followed prior work [18], where the target region of each image was center-cropped and resized to 384×384. The sample numbers for each client were 101, 159, 400, 400, and the training, validation, and test sets for each client were split in a 7:1.5:1.5 ratio. For the classification task, we used the Kaggle APTOS 2019 Blindness Detection (APTOS2019) dataset, which contains 3662 images.

Implementation Details According to the conclusion in [1], using large models on large-scale datasets for training is more advantageous. In our experiments, we

Table 2. Ablation study of the main components of FedBKD.

Method	Segmention					Classification	
	$\beta = 1/16$	$\beta = 1/8$	$\beta = 1/4$	$\beta = 1/2$	average	$\beta = 1/2$	average
	Dice	Dice	Dice	Dice	Δ_m	Acc	Δ_m
Local	79.16	79.40	88.95	90.53	0.0	71.71	0.0
FedBKD_ASE	82.36	82.79	89.87	91.00	+1.99	73.54	+1.83
FedBKD_FCO	83.21	83.92	89.36	91.44	+2.47	74.08	+2.37
FedBKD (Ours)	**83.54**	**84.14**	**90.06**	**92.21**	+2.98	**75.92**	+4.21

set K to 5 and define β as $\{1/2, 1/2, 1/4, 1/8, 1/16\}$ based on the amount of local data, where $1/2$ indicates that the client model encoder has half the capacity of the server model parameters. The global model uses RETFound [21], with Vision Transformer (ViT) [8] as the backbone. The decoder of the segmentation model uses DeepLabv3+ [5], and the decoder of the classification model uses a simple linear layer. During training, clients share only the backbone. All models are trained for 200 communication epochs. For the classification task, we use accuracy (Acc, %) as the evaluation metric, and the loss function is cross-entropy. For the segmentation task, we use Dice (%) as the evaluation metric, and the loss function is a combination of IoU and cross-entropy, with the loss function evaluation parameter λ set to 1.

3.2 Comparison with State-of-the-Arts Methods

To evaluate the performance of FedBKD, we compare it with representative studies, including traditional FL approaches, FedProx [14], and Ditto [13], as well as three methods under resource constraints: FedDrop [4], FedRolex [1], HeteroFL [7], and two FMTL methods: MaT-FL [3] and FedHCA^2 [15]. Table 1 demonstrates that FedBKD consistently outperforms all other methods across most metrics. In the segmentation task, FedBKD achieves significant improvements across all β values. Specifically, we observe an average Dice improvement of +2.98, with substantial gains in lower model capacity clients($\beta = 1/16$, $\beta = 1/8$). In comparison, methods like FedDrop [4] and FedRolex [1] show consistent performance degradation, particularly in lower-capacity models, with performance drops in both segmentation and classification tasks. For classification, FedBKD achieves an average accuracy of 75.92%, with a notable improvement of +4.21 in the average Δ_m compared to local training. The performance improvement can be attributed to FedBKD's ability to optimize feature alignment, ensuring better adaptability to task diversity and effectively demonstrating its advantage in handling diverse clients and complex tasks.

3.3 Detailed Analysis

We conduct a series of ablation experiments to evaluate the contribution of each component in FedBKD, and the experimental results are presented in Table 2. FedBKD_ASE refers to the model with the ASE module, and FedBKD_FCO

Table 3. The impact of different hyperparameters λ on the performance of segmentation and classification tasks.

Method	Segmention					Classification	
	$\beta = 1/16$	$\beta = 1/8$	$\beta = 1/4$	$\beta = 1/2$	average	$\beta = 1/2$	average
	Dice	Dice	Dice	Dice	Δ_m	Acc	Δ_m
$\lambda = 0$	82.36	82.79	89.87	91.00	0.00	73.54	0.00
$\lambda = 0.2$	83.21	82.83	89.62	91.27	+0.23	73.89	+0.35
$\lambda = 0.5$	83.47	83.54	89.90	91.23	+0.53	74.08	+0.54
$\lambda = 0.8$	83.50	83.89	89.99	91.62	+0.75	75.18	+1.64
$\lambda = 1.0$	83.54	84.14	90.06	92.21	+0.98	75.92	+2.38
$\lambda = 1.2$	81.31	82.04	89.76	90.92	-0.50	75.00	+1.46

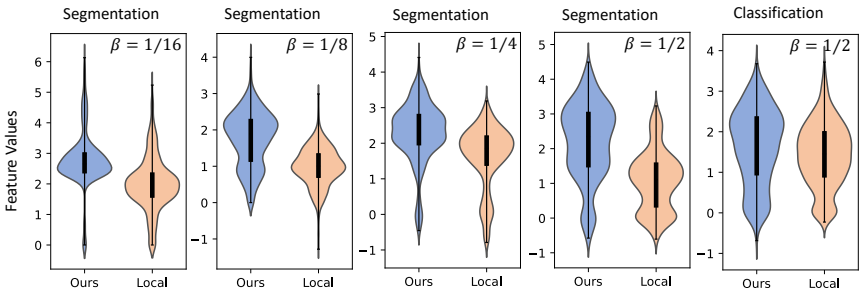

Fig. 2. The benefits of FCO via observing the feature activation value of the model encoder.

refers to the model using FCO. In segmentation tasks, FedBKD_ASE outperforms local across all β values, notably improving at $\beta = 1/16$ and $\beta = 1/8$. In classification tasks, FedBKD_ASE improves accuracy by 1.83% at $\beta = 1/2$ compared to the local training. Ultimately, FedBKD combines the strengths of both approaches, achieving the best performance with an average improvement of 2.98% in segmentation and 4.21% in classification tasks. These results demonstrate that incorporating the ASE module and FCO strategy significantly enhances the model's overall performance.

Table 3 shows that as the hyperparameter λ increases, both segmentation and classification performance improve. However, λ has a more significant impact on the classification task. This is likely because the global model only contains the shared encoder, while the segmentation task relies more on the locally specific decoder. As λ increases further to 1.2, performance slightly decreases, indicating that when λ exceeds 1, the local model becomes overly focused on aligning with the global model, neglecting the demands of the current task. Therefore, $\lambda = 1.0$ achieves the best balance between segmentation and classification tasks, providing optimal performance.

We analyze the feature distribution to assess the performance of the FCO. Suppose the model has learned effective feature representations. In that case, the relevant neurons should show strong activation [20], with activation values exceeding a predefined threshold (in this study, ReLU with a threshold of 0). The

results from five clients are shown in Fig. 2. Our analysis indicates that FedBKD achieves significantly higher feature activation values than local training. These findings demonstrate that FedBKD can learn more accurate and meaningful feature representations effectively.

4 Conclusion

In this paper, we propose FedBKD to address the challenges of FMTL in retinal image analysis. By introducing the ASE module and FCO strategy, we successfully bridge the knowledge discrepancy between the server and the client models. Under resource constraints, we effectively leverage the FM to enhance the multi-task performance of the client models. Experimental results demonstrate that FedBKD outperforms existing methods in handling task heterogeneity and resource limitations. This work offers valuable insights for future research and practical applications in medical image analysis.

Acknowledgments. This work was supported by National Natural Science Foundation of China (Grant No. 62371409) and Fujian Provincial Natural Science Foundation of China (Grant No. 2023J01005).

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

References

1. Alam, S., Liu, L., Yan, M., Zhang, M.: Fedrolex: Model-heterogeneous federated learning with rolling sub-model extraction. In: Koyejo, S., Mohamed, S., Agarwal, A., Belgrave, D., Cho, K., Oh, A. (eds.) Advances in Neural Information Processing Systems. vol. 35, pp. 29677–29690. Curran Associates, Inc. (2022)
2. Batista, F.J.F., Diaz-Aleman, T., Sigut, J., Alayon, S., Arnay, R., Angel-Pereira, D.: Rim-one dl: A unified retinal image database for assessing glaucoma using deep learning. Image Analysis and Stereology **39**(3), 161–167 (2020)
3. Cai, R., Chen, X., Liu, S., Srinivasa, J., Lee, M., Kompella, R., Wang, Z.: Many-task federated learning: A new problem setting and a simple baseline. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5037–5045 (2023)
4. Caldas, S., Konečny, J., McMahan, H.B., Talwalkar, A.: Expanding the reach of federated learning by reducing client resource requirements (2019)
5. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)
6. Cheng, G., Charles, Z., Garrett, Z., Rush, K.: Does federated dropout actually work? In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3387–3395 (June 2022)
7. Diao, E., Ding, J., Tarokh, V.: Heterofl: Computation and communication efficient federated learning for heterogeneous clients (2021)

8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
9. Golub, G.H., Van Loan, C.F.: Matrix computations. JHU press (2013)
10. Horváth, S., Laskaridis, S., Almeida, M., Leontiadis, I., Venieris, S., Lane, N.: Fjord: Fair and accurate federated learning under heterogeneous targets with ordered dropout. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems. vol. 34, pp. 12876–12889 (2021)
11. Huang, Z., Bianchi, F., Yuksekgonul, M., Montine, T.J., Zou, J.: A visual–language foundation model for pathology image analysis using medical twitter. Nature medicine **29**(9), 2307–2316 (2023)
12. Jiang, M., Le, A., Li, X., Dou, Q.: Heterogeneous personalized federated learning by local-global updates mixing via convergence rate. In: The Twelfth International Conference on Learning Representations (2024)
13. Li, T., Hu, S., Beirami, A., Smith, V.: Ditto: Fair and robust federated learning through personalization. In: International conference on machine learning. pp. 6357–6368. PMLR (2021)
14. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. Proceedings of Machine learning and systems **2**, 429–450 (2020)
15. Lu, Y., Huang, S., Yang, Y., Sirejiding, S., Ding, Y., Lu, H.: Fedhca2: Towards hetero-client federated multi-task learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5599–5609 (2024)
16. Orlando, J.I., Fu, H., Breda, J.B., Van Keer, K., Bathula, D.R., Diaz-Pinto, A., Fang, R., Heng, P.A., Kim, J., Lee, J., et al.: Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. Medical image analysis **59**, 101570 (2020)
17. Sivaswamy, J., Krishnadas, S., Chakravarty, A., Joshi, G., Tabish, A.S., et al.: A comprehensive retinal image dataset for the assessment of glaucoma from the optic nerve head analysis. JSM Biomedical Imaging Data Papers **2**(1), 1004 (2015)
18. Wang, J., Jin, Y., Stoyanov, D., Wang, L.: Feddp: Dual personalization in federated medical image segmentation. IEEE Transactions on Medical Imaging **43**(1), 297–308 (2023)
19. Wu, C., Zhang, X., Zhang, Y., Wang, Y., Xie, W.: Towards generalist foundation model for radiology. arXiv preprint arXiv:2308.02463 (2023)
20. Zhou, B., Bau, D., Oliva, A., Torralba, A.: Interpreting deep visual representations via network dissection. IEEE transactions on pattern analysis and machine intelligence **41**(9), 2131–2145 (2018)
21. Zhou, Y., Chia, M.A., Wagner, S.K., Ayhan, M.S., Williamson, D.J., Struyven, R.R., Liu, T., Xu, M., Lozano, M.G., Woodward-Court, P., et al.: A foundation model for generalizable disease detection from retinal images. Nature **622**(7981), 156–163 (2023)
22. Zhou, Z., Shen, Y., Shao, S., Gong, L., Lin, S.: Rethinking centered kernel alignment in knowledge distillation. arXiv preprint arXiv:2401.11824 (2024)