Cross-Client Gradient Alignment for Federated Multi-Source Fundus Diagnosis

Yijie Pan[1,2], Yufei Cao[3,4], Huiyan Lin[5], Haojin Li[5], Yuanchun Shi[1], Chun Yu[1], and Heng Li[3]

[1] Department of Computer Science and Technology, Tsinghua University, Beijing, China
[2] Ningbo Institute of Digital Twin, Eastern Institute of Technology, Ningbo, China
[3] Faculty of Biomedical Engineering, Shenzhen University of Advanced Technology, Shenzhen, China
[4] School of Biomedical Engineering, Shenzhen University Medical School, Shenzhen University, Shenzhen, China
[5] Department of Computer Science and Engineering, Southern University of Science and Technology, Shenzhen, China
✉ Corresponding authors: liheng@suat-sz.edu.cn

Abstract. Federated learning enables collaborative fundus diagnosis without centralizing sensitive data. However, non-IID data across clinical clients, caused by differences in imaging devices, patient populations, and acquisition protocols, often leads to inconsistent client updates and unstable aggregation. To address this issue, we propose Federated Gradient Consensus Alignment (FedGCA), a gradient-level optimization framework for heterogeneous federated diagnosis. FedGCA constructs Local Gradient Basis Memory (LGBM) to capture dominant local gradient-related directions and builds client-specific Cross-Client Gradient Consensus Matrices (CGCMs) on the server to summarize cross-client optimization trends. During local training, FedGCA uses directional consistency to detect conflicting components and adaptively suppresses them before aggregation. Experiments on five multi-source fundus datasets show that FedGCA achieves the best average AUC and F1-score among representative FL baselines. Ablation and convergence analyses further validate the effectiveness of gradient consensus alignment under multi-source data heterogeneity.

Keywords: Federated learning · Data heterogeneity · Gradient consensus alignment · Model aggregation.

1 Introduction

Federated learning (FL) has emerged as a promising paradigm for collaborative fundus diagnosis [18, 3, 5], enabling multiple clinical institutions to jointly train a global model without sharing raw patient data. In real-world clinical applications, however, participating institutions often exhibit substantial multi-source data heterogeneity due to differences in imaging devices, acquisition protocols, patient demographics, and disease prevalence [7, 22]. Such heterogeneity can lead to conflicting local updates during global aggregation, thereby degrading the convergence stability and diagnostic performance of the global model [12]. This issue is commonly reflected in reduced accuracy, slower convergence, and limited generalization across institutions.

Existing FL methods have attempted to address data heterogeneity through regularization [2], contrastive learning [15], knowledge distillation [13, 21], or adaptive aggregation strategies. Although these methods are effective to some extent, they mainly operate on model parameters, feature representations, or loss objectives, while the directional inconsistency among client gradients remains underexplored. In heterogeneous fundus diagnosis scenarios, inconsistent gradient directions may cause client updates to interfere with or cancel each other out during aggregation [11], causing the aggregated update to deviate from the desired global optimization direction. This not only slows down convergence but also weakens the contribution of clients with minority disease distributions or domain-specific imaging characteristics [22]. Explicitly addressing gradient conflicts is therefore essential for improving the stability and robustness of federated diagnostic models [25].

Mitigating gradient conflicts in heterogeneous medical FL remains challenging. i) It is difficult to identify shared gradient directions that are consistent with the global optimization objective while preserving client-specific diagnostic information. ii) Conflicting gradients need to be coordinated adaptively during training, since static constraints or history-independent adjustments may fail to capture the evolving optimization dynamics across clients. iii) Severe heterogeneity may introduce biased local updates and unstable aggregation, requiring an effective mechanism to balance global consistency with local adaptability.

To address these challenges, we propose Federated Gradient Consensus Alignment (FedGCA), a gradient-level optimization framework for heterogeneous federated fundus diagnosis. FedGCA introduces Local Gradient Basis Memory (LGBM) to capture representative gradient directions during local training, and constructs a Cross-Client Gradient Consensus Matrix (CGCM) on the server to summarize shared optimization directions across clients. During training, each client compares its local gradient direction with the received CGCM using cosine similarity. When a conflict is detected, FedGCA suppresses the conflicting component and retains the non-conflicting component for local model updating, thereby reducing inter-client gradient interference and improving aggregation stability. The main contributions of this work are summarized as follows:

- This paper proposes FedGCA, a gradient-level federated optimization framework that mitigates inter-client gradient conflicts amid heterogeneity in multi-source fundus data.
- This framework introduces Local Gradient Basis Memory and client-specific Cross-Client Gradient Consensus Matrices to model local dominant directions and cross-client optimization trends.

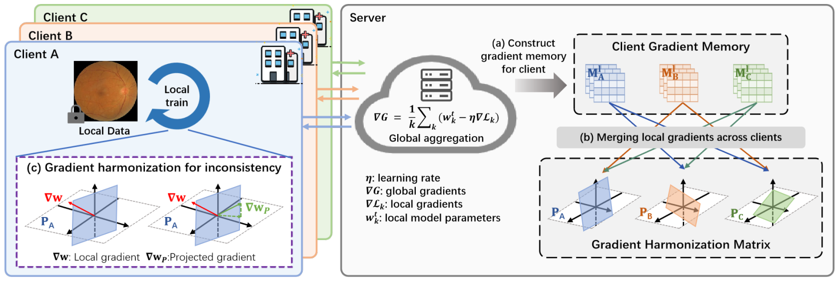

Fig. 1. Clients train diagnostic models locally and upload their layer-wise Local Gradient Basis Memories (LGBMs) to the server. (a) The server maintains the client memories $M_i = \{M_i^l\}_{l=1}^L$. (b) For each client, it constructs a layer-wise Cross-Client Gradient Consensus Matrix (CGCM) from the LGBMs of the other clients, e.g., $P_A = \{P_A^l\}_{l=1}^L$ from M_B and M_C. (c) Each client uses its received CGCM to identify directional inconsistencies and adjust the corresponding layer-wise gradients.

– FedGCA is validated on five fundus datasets and achieves the best average AUC and F1-score among representative FL baselines, while ablation studies confirm the effectiveness of gradient consensus alignment.

2 Method

As shown in Fig. 1, FedGCA consists of local gradient basis memory construction, server-side cross-client consensus construction, and consistency-guided gradient alignment. It aims to align local updates with cross-client optimization directions while suppressing conflicting gradient components.

2.1 Problem Formulation

Let K denote the number of participating clients and t the communication-round index. Client k owns a local dataset $D_k = \{(x_{k,s}, y_{k,s})\}_{s=1}^{n_k}$ containing n_k training samples. Let $w^t \in \mathbb{R}^p$ denote the global model parameters at round t, and let $\mathcal{L}_k(w)$ denote the local objective of client k. Starting from w^t, client k performs a local update as

$$w_k^{t+1} = w^t - \eta \nabla_w \mathcal{L}_k(w^t), \tag{1}$$

where η is the local learning rate. The server aggregates the locally updated models through sample-size-weighted averaging:

$$w^{t+1} = \sum_{k=1}^{K} \frac{n_k}{\sum_{j=1}^{K} n_j} w_k^{t+1}. \tag{2}$$

Under heterogeneous client distributions, local gradients may contain inconsistent or conflicting directions, making direct model averaging unstable. FedGCA mitigates this issue by identifying and suppressing conflicting gradient components before model aggregation.

2.2 Local Gradient Basis Memory

The key idea of FedGCA is to characterize dominant client-specific gradient-related directions for conflict-aware optimization. For layer l, let $x^{(l)} \in \mathbb{R}^{d_l}$ denote its input representation, $\delta^{(l)} \in \mathbb{R}^{c_l}$ the corresponding back-propagated error, and $W^{(l)} \in \mathbb{R}^{c_l \times d_l}$ the weight matrix. The layer-wise gradient can be expressed as

$$\frac{\partial \mathcal{L}}{\partial W^{(l)}} = \delta^{(l)} (x^{(l)})^\top. \tag{3}$$

For convolutional layers, the corresponding tensors are reshaped into matrix form when constructing the gradient-related subspace.

For client i, the representations from m_i local samples at layer l are arranged column-wise as $R_i^l = [x_{i,1}^{(l)}, \ldots, x_{i,m_i}^{(l)}] \in \mathbb{R}^{d_l \times m_i}$. Its column space provides a compact characterization of the corresponding gradient-related directions.

Each client constructs its Local Gradient Basis Memory (LGBM) locally by applying singular value decomposition (SVD) to R_i^l:

$$R_i^l = U_i^l \Sigma_i^l (V_i^l)^\top, \tag{4}$$

where U_i^l contains the left singular vectors and $\Sigma_i^l = \mathrm{diag}(\sigma_{i,1}^l, \ldots, \sigma_{i,q_i^l}^l)$ contains the singular values, with $q_i^l = \min(d_l, m_i)$.

The retained rank r_i^l is selected according to the cumulative energy:

$$r_i^l = \min \left\{ r : \frac{\sum_{s=1}^{r} (\sigma_{i,s}^l)^2}{\sum_{s=1}^{q_i^l} (\sigma_{i,s}^l)^2} \geq \epsilon \right\}, \tag{5}$$

where $\epsilon \in (0,1)$ is the shared energy-retention threshold. The layer-wise LGBM is then defined as

$$M_i^l = U_i^l[:, 1 : r_i^l], \qquad M_i = \{M_i^l\}_{l=1}^{L}, \tag{6}$$

where $M_i^l \in \mathbb{R}^{d_l \times r_i^l}$ and L denotes the number of considered layers. The LGBM therefore provides a compact representation of the dominant gradient-related directions of each client.

2.3 Gradient Consensus Alignment

After constructing LGBMs, the server builds client-specific Cross-Client Gradient Consensus Matrices (CGCMs) by integrating the LGBMs of other clients. Each client then uses the received CGCM to detect directional conflicts and adjust local gradients before aggregation.

Cross-Client Gradient Consensus Matrix. For client C_j, the server constructs a layer-wise CGCM $\mathbf{P}_j = \{\mathbf{P}_j^l\}_{l=1}^{L}$ exclusively from the LGBMs of the

Table 1. Class distributions of the five fundus datasets.

Datasets	AMD	DR	Glaucoma	Myopia	Normal	Cataract
ISEE [14]	666	231	394	728	7770	-
ADAM [4]	89	-	-	-	311	-
Retina [10]	-	-	101	-	300	100
ODIR [1]	235	1462	228	237	3106	276
MuReD [20]	158	495	-	89	490	-

other clients. For each layer l, $\mathbf{P}_j^l$ is initialized as empty and sequentially updated using M_i^l from each client $i \neq j$. To remove directions already represented by the current consensus basis, we compute

$$\widetilde{M}_{i\to j}^{l} = \left(I_{d_l} - \mathbf{P}_j^l(\mathbf{P}_j^l)^\top\right) M_i^l. \tag{7}$$

SVD is then applied to $\widetilde{M}_{i\to j}^{l}$, and the smallest number of dominant singular vectors whose cumulative energy reaches ϵ is retained. These non-redundant directions are appended to the CGCM:

$$\mathbf{P}_j^l = \left[\mathbf{P}_j^l,\, \widetilde{U}_{i\to j}^{l}[:, 1 : a_{i\to j}^l]\right], \tag{8}$$

where $a_{i\to j}^l$ denotes the number of newly retained directions. Repeating this procedure over all $i \neq j$ yields the client-specific CGCM.

Consistency-Guided Gradient Alignment for Aggregation. After receiving its CGCM, each client evaluates the directional consistency between its local gradient-related directions and the corresponding cross-client consensus directions layer by layer. Let S_i^l denote the layer-wise consistency score, computed by cosine similarity between the corresponding local and cross-client gradient-related directions. When $S_i^l \leq 0$, the local update is regarded as conflicting with the consensus subspace and its conflicting component is removed:

$$\widehat{G}_i^l = \begin{cases} G_i^l\left(I_{d_l} - \mathbf{P}_i^l(\mathbf{P}_i^l)^\top\right), & S_i^l \leq 0, \\ G_i^l, & S_i^l > 0, \end{cases} \tag{9}$$

where G_i^l denotes the reshaped gradient matrix of layer l and I_{d_l} is the d_l-dimensional identity matrix. The adjusted gradient is restored to its original tensor shape for local optimization, after which the updated client models are aggregated according to Eq. 2.

3 Experiments

3.1 Datasets and Implementation Details

We evaluate FedGCA on five public fundus datasets, including ISEE [14], ADAM [4], Retina [10], ODIR [1], and MuReD [20], treating each dataset as one client. The task is six-category multi-class classification with a shared global label space and prediction head. A category absent from a client simply has no local samples, while its label definition remains unchanged, resulting in heterogeneous class distributions across clients.

All fundus images are resized to 224×224 and normalized before being fed into the network. ResNet-18 [6] is adopted as the backbone for all compared methods. The models are trained using SGD with a learning rate of 0.001 and a batch size of 32. Each communication round contains one local training epoch, and the global model is trained for 150 rounds. The model with the best validation performance is selected for testing. For FedGCA, the shared energy-retention threshold ϵ is set to 0.95 for all clients and layers, retaining 95% of the representation energy while keeping the gradient basis compact. All experiments are implemented with PyTorch and conducted under the same training protocol for fair comparison.

Table 2. Performance of global models on five fundus clients. Avg. denotes mean±std across clients.

AUC						
Model	ISEE	ADAM	Retina	ODIR	MuReD	Avg.
FedAvg [17]	67.22	54.32	75.00	69.68	55.49	64.34±8.10
FedProx [16]	69.21	59.55	66.67	68.73	49.90	62.81±7.32
FedDyn [2]	49.27	81.25	47.19	57.31	75.43	62.09±13.82
FedCGP [19]	66.19	65.45	83.33	69.86	66.02	70.17±6.76
FedNTD [13]	71.49	47.95	80.83	74.23	58.63	66.63±11.79
FPL [9]	68.35	61.82	87.50	65.88	76.10	71.93±9.07
FedDisco [24]	69.42	61.36	43.33	75.26	65.57	62.99±10.84
FedCross [8]	62.58	72.46	84.74	68.07	70.52	71.67±7.33
FedAS [23]	55.49	62.00	59.26	62.68	61.66	60.22±2.63
FedGCA	63.95	67.95	85.83	71.64	71.28	**72.13±7.39**
F1-score						
Model	ISEE	ADAM	Retina	ODIR	MuReD	Avg.
FedAvg [17]	56.25	54.48	62.50	45.92	7.69	45.37±20.01
FedProx [16]	55.31	38.78	26.79	41.34	4.23	33.29±17.13
FedDyn [2]	15.12	65.89	17.03	28.94	56.81	36.76±20.83
FedCGP [19]	74.82	56.02	90.73	44.20	8.84	54.92±27.99
FedNTD [13]	72.00	56.02	82.24	42.80	6.60	51.93±26.37
FPL [9]	75.33	56.02	89.14	44.99	9.98	55.10±27.64
FedDisco [24]	75.84	56.02	90.73	44.64	8.22	55.09±28.37
FedCross [8]	72.64	76.56	51.49	45.04	33.58	55.86±16.39
FedAS [23]	70.79	64.29	47.08	40.27	20.96	48.68±17.74
FedGCA	74.57	68.85	89.14	50.76	41.54	**64.97±16.96**

3.2 Comparison with State-of-the-art FL Methods

Table 2 reports the performance of different FL methods on multi-source fundus image diagnosis. FedGCA achieves the best overall performance, with an average AUC of 72.13±7.39 and an average F1-score of 64.97±16.96. Compared with FedAvg, FedGCA substantially improves the average AUC and F1-score, although the gain varies across clients.

Among the compared methods, FedProx and FedDyn alleviate client drift through regularization-based mechanisms, but their improvements are limited under severe cross-client distribution shifts. FedNTD and FPL improve performance by introducing knowledge distillation or prototype-based learning, while FedDisco and FedCross enhance aggregation under heterogeneous settings. However, these methods still show lower average F1-scores than FedGCA, suggesting that gradient-level conflicts among clients may not be fully mitigated. In contrast, FedGCA directly aligns client update directions in the gradient space, leading to more stable global aggregation and better diagnostic performance.

The advantage of FedGCA is particularly evident in F1-score. This suggests that FedGCA is beneficial for heterogeneous and imbalanced fundus diagnosis scenarios, where minority disease categories or smaller clients are more likely to be affected by conflicting updates. By suppressing gradient components that conflict with cross-client consensus directions, FedGCA better preserves useful local diagnostic information while maintaining global optimization consistency.

3.3 Visualization and Convergence Analysis

To further investigate the diagnostic behavior of different methods, Figure 2 presents CAM visualizations on representative fundus images. Clinically, myopia and glaucoma are closely associated with abnormalities around the optic

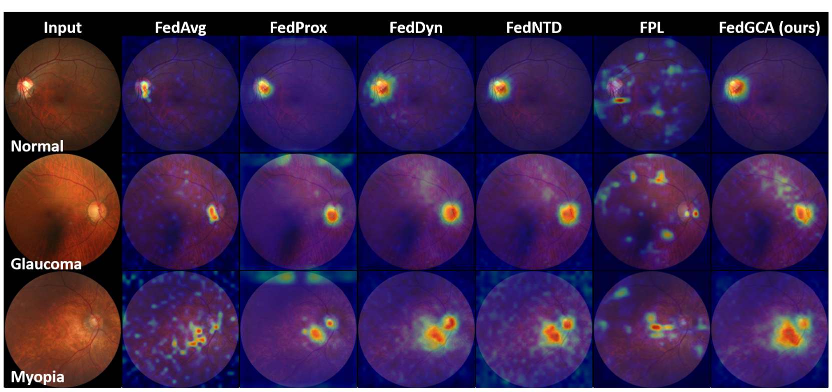

Fig. 2. CAM visualizations of representative fundus images. Compared with baseline methods, FedGCA produces more concentrated activation responses around clinically relevant retinal regions.

disc and surrounding retinal structures. Compared with FedAvg, FedProx, FedDyn, FedNTD, and FPL, FedGCA focuses more accurately on disease-related regions and produces more concentrated activation responses around clinically meaningful areas. This indicates that gradient consensus alignment helps global model learn more discriminative and disease-relevant representations from heterogeneous clients.

Figure 3 shows the AUC curves of FedAvg, FedNTD, and FedGCA during training on fundus datasets. FedAvg exhibits large fluctuations across different clients, suggesting that simple parameter averaging is sensitive to cross-client heterogeneity. FedNTD improves the stability to some extent, but its curves still show noticeable performance gaps among clients. In contrast, FedGCA achieves smoother and more consistent AUC trends across datasets. This demonstrates that the proposed gradient consensus alignment mechanism can reduce inter-client update conflicts, stabilize global optimization, and improve convergence behavior in multi-source federated diagnosis.

3.4 Ablation Study

Table 3 presents the ablation study on multi-source fundus diagnosis. Starting from the FedAvg baseline, introducing gradient consensus alignment improves the average AUC from 64.34 to 70.32 and the average F1-score from 45.37 to 57.75. This confirms that removing conflicting gradient components can effectively mitigate cross-client interference and improve server aggregation.

When the consistency constraint is further incorporated, the performance increases to 72.13 in AUC and 64.97 in F1-score. This shows that directional consistency provides an effective criterion for deciding when gradient adjustment should be applied, helping FedGCA balance global consistency and local adaptability.

FedGCA introduces additional computation for SVD and gradient projection and communication for transmitting compact LGBMs, whose size is controlled by ϵ. Although these bases may still encode client-specific information without a formal privacy guarantee, future work will investigate privacy-preserving mechanisms, alternative backbones, and partial client participation.

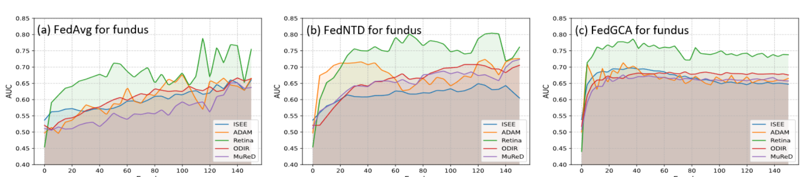

Fig. 3. AUC curves of FedAvg, FedNTD, and FedGCA during training on five fundus clients. FedGCA shows smoother and more consistent convergence trends under multi-source heterogeneity.

Table 3. Ablation study on fundus datasets.

Baseline	GCA	Consistency	AUC	F1
✓			64.34	45.37
✓	✓		70.32	57.75
✓	✓	✓	**72.13**	**64.97**

4 Conclusion

In this paper, we proposed FedGCA, a gradient-level consensus alignment framework for federated fundus diagnosis under multi-source data heterogeneity. FedGCA uses Local Gradient Basis Memory to capture dominant local directions and Cross-Client Gradient Consensus Matrices to summarize cross-client optimization trends. By suppressing conflicting gradient components based on directional consistency, FedGCA improves aggregation stability and diagnostic performance. Experiments on five fundus datasets show that FedGCA outperforms representative FL baselines on average, and ablation studies further confirm the effectiveness of its key components.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62401246, the Guangdong Basic and Applied Basic Research Foundation under Grant No. 2025A1515010826, and the Shenzhen Science and Technology Program under Grant Nos. JCYJ20250604185805008 and JCYJ20240813095112017. The authors gratefully acknowledge Beijing Tiromu Medical Technology Co., Ltd. and Shenzhen DE Sci&Tech Co., Ltd. for their support in technical validation and for providing a practical platform for this work.

References

1. Ocular Disease Intelligent Recognition ODIR-5K. https://academictorrents.com/details/cf3b8d5ecdd4284eb9b3a80fcfe9b1d621548f72 (2019), dataset, Academic Torrents, accessed 26 May 2026
2. Acar, D.A.E., Zhao, Y., Navarro, R.M., Mattina, M., Whatmough, P.N., Saligrama, V.: Federated learning based on dynamic regularization. In: Proceedings of the International Conference on Learning Representations (ICLR) (2021)
3. Bali, A., Mansotra, V.: Analysis of deep learning techniques for prediction of eye diseases: A systematic review. Archives of Computational Methods in Engineering **31**(1), 487–520 (2024)
4. Fu, H., Li, F., Orlando, J.I., Bogunović, H., Sun, X., Liao, J., Xu, Y., Zhang, S., Zhang, X.: Adam: Automatic detection challenge on age-related macular degeneration. IEEE Dataport (2020), [Online]. Available: http://dx.doi.org/10.21227/dt4f-rt59

5. Hanif, A., Lu, C., Chang, K., Singh, P., Coyner, A.S., Brown, J.M., Ostmo, S., Chan, R.V.P., Rubin, D., Chiang, M.F., et al.: Federated learning for multicenter collaboration in ophthalmology: implications for clinical diagnosis and disease epidemiology. Ophthalmology retina **6**(8), 650–656 (2022)
6. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016)
7. He, T., Han, P., Duan, S., Wang, Z., Wu, W., Liu, C., Han, J.: Generative data augmentation with differential privacy for non-iid problem in decentralized clinical machine learning. Future Generation Computer Systems **160**, 171–184 (2024)
8. Hu, M., Zhou, P., Yue, Z., Ling, Z., Huang, Y., Li, A., Liu, Y., Lian, X., Chen, M.: Fedcross: Towards accurate federated learning via multi-model cross-aggregation. In: 2024 IEEE 40th International Conference on Data Engineering (ICDE). pp. 2137–2150. IEEE (2024)
9. Huang, W., Ye, M., Shi, Z., Li, H., Du, B.: Rethinking federated learning with domain shift: A prototype view. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16312–16322 (2023)
10. jr2ngb: Cataract Dataset. https://www.kaggle.com/datasets/jr2ngb/cataractdataset, kaggle dataset, accessed 26 May 2026
11. Kairouz, P., McMahan, H.B.: Advances and open problems in federated learning. Foundations and trends in machine learning **14**(1-2), 1–210 (2021)
12. Karimireddy, S.P., Kale, S., Mohri, M., Reddi, S.J., Stich, S.U., Suresh, A.T.: Scaffold: Stochastic controlled averaging for federated learning. In: Proceedings of the International Conference on Machine Learning (ICML). vol. 119, pp. 5132–5143 (2020)
13. Lee, G., Jeong, M., Shin, Y., Bae, S., Yun, S.Y.: Preservation of the global knowledge by not-true distillation in federated learning. In: Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS). vol. 35, pp. 38461–38474 (2022)
14. Li, H., Liu, H., Hu, Y., Fu, H., Zhao, Y., Miao, H., Liu, J.: An annotation-free restoration network for cataractous fundus images. IEEE Transactions on Medical Imaging **41**(7), 1699–1710 (2022)
15. Li, Q., He, B., Song, D.: Model-contrastive federated learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10713–10722 (2021)
16. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated optimization in heterogeneous networks. In: Proceedings of the Conference on Machine Learning and Systems (MLSys). vol. 2, pp. 429–450 (2020)
17. McMahan, H.B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS). vol. 54, pp. 1273–1282 (2017)
18. Nguyen, T.X., Ran, A.R., Hu, X., Yang, D., Jiang, M., Dou, Q., Cheung, C.Y.: Federated learning in ocular imaging: current progress and future direction. Diagnostics **12**(11), 2835 (2022)
19. Niu, K., Tai, W., Cai, J., Zhou, Y., Li, H.: Fedcgp: Cluster-based gradual personalization for federated medical image segmentation. IEEE Transactions on Emerging Topics in Computational Intelligence (2025)
20. Rodríguez, M.A., AlMarzouqi, H., Liatsis, P.: Multi-label retinal disease classification using transformers. IEEE Journal of Biomedical and Health Informatics **27**(6), 2739–2750 (2022)

21. Salman, H., Zaki, C., Charara, N., Guehis, S., Pradat-Peyre, J.F., Nasser, A.: Knowledge distillation in federated learning: a comprehensive survey. Discover Computing **28**(1), 145 (2025)
22. Wang, Y., Fu, H., Kanagavelu, R., Wei, Q., Liu, Y., Goh, R.S.M.: An aggregation-free federated learning for tackling data heterogeneity. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 26233–26242 (2024)
23. Yang, X., Huang, W., Ye, M.: Fedas: Bridging inconsistency in personalized federated learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11986–11995 (2024)
24. Ye, R., Xu, M., Wang, J., Xu, C., Chen, S., Wang, Y.: Feddisco: Federated learning with discrepancy-aware collaboration. In: Proceedings of the International Conference on Machine Learning (ICML). pp. 39879–39902 (2023)
25. Yu, T., Kumar, S., Gupta, A., Levine, S., Hausman, K., Finn, C.: Gradient surgery for multi-task learning. In: Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS). vol. 33, pp. 5824–5836 (2020)