Shuffle-Diversity Collaborative Federated Learning for Imbalanced Medical Image Analysis

Wenpeng Gao[1], Liantao Lan[1], Yumeng Liu[2], Ruxin Wang[3(✉)], and Xiaomao Fan[2(✉)]

[1] South China Agricultural University
gaowp@stu.scau.edu.cn; lanlt@scau.edu.cn
[2] Shenzhen Technology University
liuyumeng@sztu.edu.cn; astrofan2008@gmail.com
[3] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences
rx.wang@siat.ac.cn

Abstract. Data imbalance presents a significant challenge for the application of federated learning in medical image analysis. To address this challenge, we propose FedSDC, an innovative federated approach designed to effectively tackle the issue of data imbalance, as well as heterogeneity in distributed federated learning environments. The proposed FedSDC framework comprises a shared body network and multiple task-specific head networks. By incorporating a shuffle-diversity collaborative strategy, FedSDC effectively addresses data imbalanc and heterogeneity challenges while improving cross-client generalization. Furthermore, training multiple heads under this strategy enables ensemble predictions, which enhances decision stability and accuracy. To balance efficiency and performance, FedSDC employs the sparse-head scheme during inference phase. Extensive experiments on medical image classification tasks validate that FedSDC achieves state-of-the-art results under imbalanced and heterogeneous data conditions. The source code will be available at `https://github.com/wpnine/FedSDC`.

Keywords: Federated Learning · Medical Image Classification · Heterogeneity.

1 Introduction

The emergence and rapid development of deep learning have revolutionized the field of medical image analysis [14,3]. To ensure a deep neural network maintains reliable performance and generalization when applied to diverse clinical centers, an extensive collection of medical image datasets from multiple sources is needed. However, in real-world scenarios, due to patient privacy and legal regulatory policy to data sharing, it is difficult to integrate patient data from multiple medical

* Wenpeng Gao and Liantao Lan contribute equally to this work. Xiaomao Fan and Ruxin Wang are co-corresponding authors.

institutions. Different conventional centralized learning systems, federated learning (FL), has been proposed as a promising alternative technique for accessing large-scale data information and training deep neural networks across hospitals without accessing raw data [16,26,10].

Despite FL-based technology's promising potential, a key challenge impeding the further development of FL is data imbalance and heterogeneity [18,5] from various medical centers, i.e., data with different distributions, formats, or structures across multiple sources or devices. For medical image data, imbalanced label distribution (even missing some categories across hospitals) and image heterogeneity are common issues [27]. These issues lead to different local models being optimized toward distinct local objectives, resulting in divergent optimization directions [25]. Consequently, aggregating these divergent local models to obtain a robust global model becomes challenging. To empirically demonstrate this issue, we conduct controlled experiments using the Matek-19 [17] dataset divided into three clients, comparing model performance under both balanced and artificially imbalanced conditions (created by selectively limiting three categories to simulate real-world data skew). As shown in Figure 1, conventional approaches like FedRep [6] exhibit substantial performance degradation when faced with inter-client categorical imbalances. This empirical evidence underscores the pressing need for adaptive strategies that can handle heterogeneous data distributions in practical deployment scenarios.

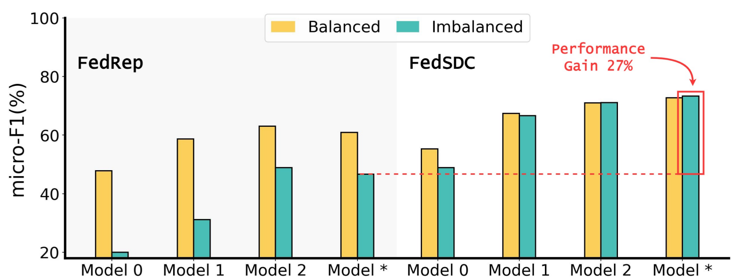


Fig. 1. The experimental results compare the federated learning methods of FedRep and our proposed FedSDC on the Matek-19 dataset, divided among three clients.

Solutions to address data imbalance and heterogeneity in federated learning can be broadly categorized into generalized federated learning (GFL) and personalized federated learning (PFL) [7]. The former aims to construct a global model, mitigating personal differences by imposing constraints on local training [11,12,29,20], modifying logits [13,30], adjusting the weights of submitted gradients [24], or generating synthetic data [32,15]. In contrast, PFL focuses on tailoring local models to adapt to the data or system of each client, placing relatively less emphasis on locally missing classes and selectively sharing either partial network parameters [6,2] or class prototypes [21] to minimize the impact of personal characteristics [8]. In the context of automated medical image classification, GFL is an ideal general solution. However, due to the severe impact of data heterogeneity, the aggregated global model may fail to achieve the performance expected of the local models. On the other hand, PFL can better fit the local data of all participating institutions, but the overall generalization capability of individual local models is often inferior to that of a global model. Both approaches have their advantages and disadvantages. *Could the strengths of these two methods be further combined to propose a federated learning approach better suited for the task of automated medical image classification task?*

In this paper, a novel heterogeneous federated learning method, FedSDC, is proposed with a body and multiple heads for medical image classification tasks. The Body is responsible for learning global data representations and serves as a shared feature extractor following the standard GFL paradigm. In contrast, the heads are designed to retain personalized information specific to each client's data and act as decision-makers for the final outputs. The training process for the heads follows a designed shuffle-diversity collaborative strategy to promote generalization performance across clients for imbalanced data distribution. Meantime, each head contributes to ensemble learning and enables more reliable decision-making. The proposed FedSDC effectively blends the strengths of both GFL and PFL, which combines both generalization and personalization. The results demonstrate that FedSDC achieves outstanding performance under both heterogeneous and non-heterogeneous data scenarios.

2 Methodology

2.1 Preliminaries

The standard formulation of federated learning involving N clients is expressed as follows:

$$\min_{(m_1,\ldots,m_N)\in\varrho_N} \frac{1}{N}\sum_{i=1}^{N} f_i(m_i), \tag{1}$$

where f_i represents the error function, while m_i denotes the learning model associated with the i-th client. The set ϱ_N signifies the feasible space of N models. This framework operates within a supervised learning context, where the data for each client i is derived from a distribution denoted as $(\mathbf{x}_j, y_j) \sim \mathcal{D}_i$. The learning model m_i is designed to map the input features $\mathbf{x}_j$ to predicted outcomes $m_i(\mathbf{x}_j) \in \mathcal{Y}$, which should ideally align with the true labels y_j. The error function f_i is defined as the expected risk over the distribution $\mathcal{D}_i$:

$$f_i\left(m_i\right) := \mathbb{E}_{(\mathbf{x}_j,y_j)\sim\mathcal{D}_i}\left[\ell\left(m_i\left(\mathbf{x}_j\right), y_j\right)\right]. \tag{2}$$

Here, ℓ serves as the loss function that quantifies the discrepancy between the predicted label $m_i(\mathbf{x}_j)$ and the actual label y_j.

2.2 Network Architecture

Figure 2 illustrates the architecture of FedSDC, a federated learning framework designed to address categorical imbalance across clients while enhancing model

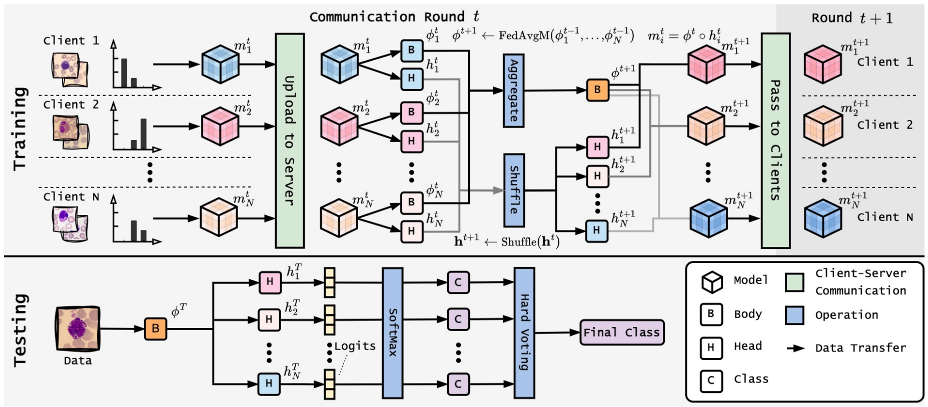


Fig. 2. The overall architecture of FedSDC. During training, the server partitions the model parameters uploaded by clients into two components: the body and the head. It aggregates multiple bodies into a shared body (ϕ), shuffles the heads, and combines each head with ϕ to form a complete model for each head. During testing, FedSDC utilizes ϕ to extract features, which are then passed to the heads for ensemble prediction with a major voting scheme.

robustness. The framework partitions each client's model into two components: a shared body network and a client-specific heterogeneous head. The shared body ϕ employs a ResNet50 backbone to learn generalized cross-client feature representations. To further amplify head heterogeneity, a Diversity mechanism is incorporated into the head architecture: each heterogeneous head h_i consists of a two-layer fully connected (FC) network, where a dropout layer is inserted between the linear layers to mitigate overfitting and promote structural diversity, and the dimensionality of the first FC layer in h_i is dynamically adjusted via a client-specific compression factor $\rho_i \in [0.4, 1]$. This configuration not only diversifies client-specific decision boundaries but also optimizes compatibility with ensemble-based inference, improving overall testing-phase performance.

Specifically, at communication round t in the training phase (see Algorithm 1), all client models $\{m_i^t\}_{i=1}^N$ are uploaded to the center server. Each client model m_i^t undergoes dual-component decomposition:

$$m_i^t \to \{\phi_i^t, h_i^t\}, \tag{3}$$

where ϕ_i^t denotes the body network and h_i^t represents the head network. In order to enhance robustness and accelerate convergence under non-IID conditions, the body networks $\phi_{i\,i=1}^{t\,N}$ are aggregated using FedAvgM [9]:

$$v^t \leftarrow \beta v^{t-1} + \textstyle\sum_{i=1}^N \frac{|\mathcal{D}_i|}{\sum_{j=1}^N |\mathcal{D}_j|} \Delta\phi_i^t, \tag{4}$$

$$\phi^{t+1} \leftarrow \phi^t - v^t, \tag{5}$$

where $\Delta\phi_i^t$ is the weight of body update from the i-th client in t-th round and $\beta \in [0,1]$ is the momentum parameter, which is empirically set to 0.5. Head networks $\{h_i\}_{i=1}^N$ undergo client-wise permutation:

$$\{h_i^{t+1}\}_{i=1}^N = \text{Shuffle}(\{h_1^t, h_2^t, \cdots, h_N^t\}). \tag{6}$$

This randomized head reassignment strategy, termed Shuffle, is designed to promote exposure to diverse decision boundaries across clients. Reconstructed client models combine the updated body network with permuted heads:

$$m_i^{t+1} \leftarrow \phi^t \circ h_i^t, \tag{7}$$

where $\circ$ denotes the functional composition. Each client subsequently optimize its local model through:

$$\min_{m_i} \mathbb{E}_{(x_j,y_j)\sim\mathcal{D}_i}\Big[\ell(h_i(\phi_i(x_j)), y_j)\Big], \tag{8}$$

Using Stochastic Gradient Descend (SGD) with learning rate η, where $\ell(\cdot)$ represents the cross-entropy function. During the testing phase, FedSDC employs an ensemble framework comprising a shared body network ϕ^T and multiple client-specific heads $\{h_i^T\}_{i=1}^N$. The input x is first encoded into a generalized feature representation via ϕ^T, which is then processed by all heads to generate head-specific class probabilities $p_i = \text{Softmax}(h_i^T(\phi^T(x)))$; these predictions $\{p_i\}_{i=1}^N$ are aggregated via major voting to produce the final ouput $\hat{y}$. To optimize efficiency and performance, a sparse subset of heads is dynamically selected for testing based on validation-phase micro-F1 scores, retaining only the top $\gamma \in [0,1]$ fraction (e.g., $\gamma = 0.3$ prunes 70% under-performing heads). Specifically, each head is evaluated on the target dataset to obtain a micro-F1 score, and only the top-performing heads are retained for ensemble prediction, thereby reducing computational overhead while preserving ensemble diversity and prediction quality. In this paper, the FedSDC with sparse heads is termed FedSDC$^+$.

3 Experiment

3.1 Datasets

In our experiments, we introduce three White Blood Cell datasets (Matek-19 [17], Acevedo-20 [1], and Bodzas-23 [4]) and one skin cancer dataset (HAM10000 [22]). We construct two federated scenarios to evaluate the performance:

IID: As shown in Figure 3-a, **IID** employs Acevedo-20 dataset that contains independent and identically distributed (i.i.d.) data with identical label distribution across clients. This dataset is split into 80% for training and 20% for testing, with the training data evenly distributed across clients.

NIID: For non-i.i.d. data with imbalanced label distribution scenarios, **NIID-1**(Figure 3-b) is created using HAM10000 dataset. The HAM10000 dataset is sourced from two different regions and exhibits a significant imbalance. We follow the data distribution setup in [28], where 30 samples per class are reserved for

Algorithm 1 FedSDC

Input: N: the total number of clients, T: communication rounds, $[\mathcal{D}_1, \ldots, \mathcal{D}_N]$: dataset of each client

Server executes:
Initialize $\phi^1, \mathbf{h}^1 = [h_1^1, \ldots, h_N^1]$
for round $t = 1$ **to** T **do**
  for client $i = 1$ **to** N **in parallel do**
    $m_i^t \leftarrow \text{ClientUpdate}(m_i^t)$ `// model sends to client for update`
    $m_i^t \rightarrow \{\phi_i^t, h_i^t\}$ `// according to Eq.(3)`
  end
  $\phi^{(t+1)} \leftarrow \text{FedAvgM}(\{\phi_1^t, \ldots, \phi_N^t\})$ `// according to Eq.(4)`
  $\{h_i^{t+1}\}_{i=1}^N = \text{Shuffle}(\{h_1^t, \cdots, h_N^t\})$ `// according to Eq.(6)`
  $\{m_1^{t+1}, \ldots, m_N^{t+1}\} \leftarrow \{\phi^{t+1} \circ h_1^{t+1}, \ldots, \phi^{t+1} \circ h_N^{t+1}\}$ `// according to Eq.(7)`
end
ClientUpdate(m_i):
Initialize client model with m_i
$m_i \leftarrow \text{SGD}(\mathcal{D}_i, m_i)$
return m_i **to** the center server

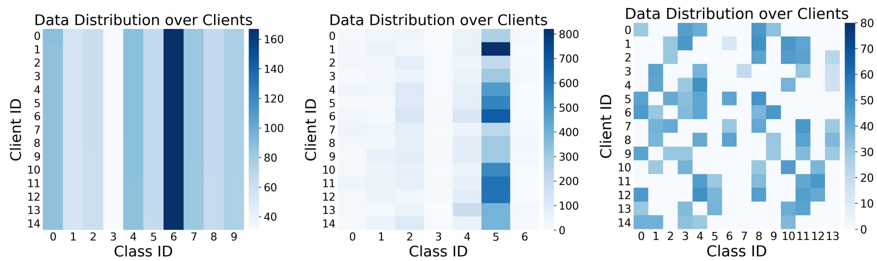
testing, and the remaining training data is distributed using a Dirichlet distribution [23] with $\beta = 5.0$; **NIID-2** (Figure 3-c) includes the Matek-19, Acevedo-20, and Bodzas-23 datasets, which simulates a more heterogeneous real-world scenario under data heterogeneity and imbalanced label distribution. These datasets inherently exhibit non-i.i.d. characteristics, including missing classes. For the training set, each dataset is split into 5 client data distributions, with each client randomly selecting 5 to 6 categories [28], and the sample size for each category randomly ranges from 20 to 60. The remaining samples are used for testing.

3.2 Baselines

In this paper, we evaluate seven baselines for comparison, including: (1) **FedAvgM** [9], (2) **FedAdam** [19], (3) **FedYogi** [19], and (4) **FedAdagrad** [19], which are standard FL aggregation algorithms designed to address data heterogeneity; (5) **FedProx** [12], which mitigates heterogeneity challenges during local training by introducing a proximal term; (6) **FedDisco** [28] and (7) **ISFL** [31], two state-of-the-art methods for heterogeneous federated learning.

3.3 Implementation Details

All the methods are implemented using PyTorch 2.4.1 with four NVIDIA 4090 GPUs. The maximum communication rounds for the server are set to 1500, and early stopping is supported. To optimize the settings for different data, for the IID and NIID-2, we set the optimizer to SGD with a learning rate of 1e-4, and the batch size is set to 32. For the NIID-1, the learning rate is set to 1e-3 and



(a) IID(10-class) (b) NIID-1(7-class) (c) NIID-2(14-class)

Fig. 3. Data distribution across clients. **IID** only includes 11,630 training samples, where all client samples are evenly distributed. **NIID-1** contains 9,805 training samples, exhibiting a significant imbalance in distribution. **NIID-2** includes a total of 3,115 training samples, presenting non-i.i.d. issues and imbalanced label distribution.

trained with a batch size of 64. All methods are run five times with different random seeds and report the average micro-F1 results along with the variance.

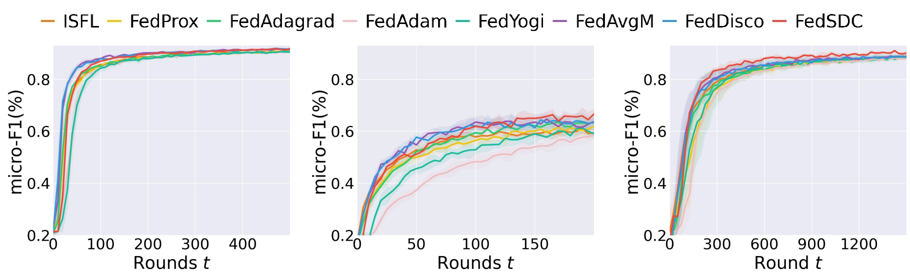
3.4 Performance Comparison

Table 1 compares the micro-F1 scores of FL methods across three datasets: IID, NIID-1, and NIID-2. Under IID conditions, all methods achieve strong performance, with FedAvgM scoring 91.97% and FedSDC slightly outperforming it at 92.01%, underscoring its efficacy in homogeneous data environments. This aligns with expectations, as uniform data distribution inherently simplifies FL optimization. In contrast, NIID settings reveal stark performance variations. For NIID-1, all baselines exhibit significant degradation, with FedSDC scoring 68.76%—highlighting the challenges of client-specific data skew. Notably, FedSDC$^+$ improves NIID-1 performance to 70.00%, demonstrating the value of its algorithmic enhancements. Meanwhile, FedSDC regains robustness in NIID-2, achieving 91.37%, which suggests tailored adjustments can mitigate NIID effects. Collectively, these results emphasize the necessity of addressing data heterogeneity in FL and position FedSDC and its variants as promising frameworks for medical imaging, where data distributions are often inherently non-uniform.

Figure 4 illustrates the convergence behaviors of FL models across IID, NIID-1, and NIID-2 settings. In the IID scenario, all models attain rapid convergence to high micro-F1 scores (>90%), reflecting their suitability for uniform data. Under NIID-1, however, convergence patterns diverge sharply: FedSDC shows markedly slower progress, mirroring the performance drop in Table 1, while FedSDC$^+$ achieves steadier improvement. The NIID-2 setting reveals enhanced stability, with most models converging more smoothly—a trend likely attributable to architectural adaptations for heterogeneity. These curves underscore the critical role of algorithmic resilience in non-i.i.d. FL, particularly in applications like medical imaging where data variability is pervasive.

Table 1. Experimental results with the evaluation metric of micro-F1 Score

Method	**IID(%)**	**NIID-1(%)**	**NIID-2(%)**
FedAvgM [9]	91.97 ± 0.26	66.29 ± 1.63	89.52 ± 0.65
FedAdam [19]	90.78 ± 0.21	58.95 ± 1.89	88.52 ± 0.59
FedYogi [19]	90.85 ± 0.36	62.00 ± 2.06	89.15 ± 1.09
FedAdagrad [19]	91.06 ± 0.51	65.24 ± 1.39	88.82 ± 0.51
FedProx [12]	90.96 ± 0.12	63.71 ± 1.63	89.15 ± 1.09
FedDisco [28]	91.95 ± 0.23	65.91 ± 2.64	89.39 ± 0.76
ISFL [31]	91.00 ± 0.24	61.33 ± 1.48	89.18 ± 0.67
FedSDC	**92.01 ± 0.26**	68.76 ± 0.99	91.37 ± 0.25
FedSDC^{+}	91.91 ± 0.21	**70.00 ± 1.12**	**91.41 ± 0.25**



(a) IID (b) NIID-1 (c) NIID-2

Fig. 4. Convergence curve plots of models under different scenario settings.

3.5 Ablation Study

Table 2 presents the ablation study results conducted in the NIID-1 dataset, evaluating the performance of the model with various strategies such as Diversity, Shuffle, and Ensemble. Model M1, which lacks both Diversity and Shuffle, scored 56.67%, indicating limited effectiveness. Introducing the Shuffle feature in Model M2 leads to a slight improvement to 57.62%. Model M3, incorporating both Diversity and Shuffle features, achieves a significant enhancement with a score of 66.29%. Finally, FedSDC obtains the highest micro-F1 score of 68.76% with the proposed shuffle-diversity collaborative strategy. These results highlight the importance of training multiple heads under this strategy, which enables the best performance.

Table 2. Experimental results of the ablation study in NIID-1 scenario

Model	Diversity	Shuffle	micro-F1(%)
M1	✗	✗	56.67±0.96
M2	✓	✗	57.62±1.03
M3	✗	✓	66.29±1.70
FedSDC	✓	✓	**68.76±0.99**

4 Conclusion

In this paper, we propose a novel generalized federated learning method, namely FedSDC, designed to address the issue of imbalanced data distribution in medical image analysis. By synergizing feature sharing (body) with structured heterogeneity (heads), our framework overcomes the limitations of conventional aggregation strategies. The introduced sparse head mechanism (FedSDC^{+}) further reduces computational overhead while preserving ensemble diversity and prediction quality. Experiments on three datasets verify the efficacy of our FedSDC and FedSDC^{+}.

Acknowledgments. This work was supported by the Special subject on Agriculture and Social Development, Key Research and Development Plan in Guangzhou (No.2023B03J0172), the National Natural Science Foundation of China (No. 12471308), the Science and Technology Program of Guangzhou (No. 2024A04J6542), and the Guangdong Basic and Applied Basic Research Foundation (No.2022B1515130009).

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

References

1. Acevedo, A., Merino, A., Alférez, S., Ángel Molina, Boldú, L., Rodellar, J.: A dataset of microscopic peripheral blood cell images for development of automatic recognition systems. Data in Brief **30**, 105474 (2020)
2. Arivazhagan, M.G., Aggarwal, V., Singh, A.K., Choudhary, S.: Federated Learning with Personalization Layers. arXiv preprint arXiv:1912.00818 (2019)
3. Barragán-Montero, A., Javaid, U., Valdés, G., Nguyen, D., Desbordes, P., Macq, B., Willems, S., Vandewinckele, L., Holmström, M., Löfman, F., et al.: Artificial Intelligence and Machine Learning for Medical Imaging: A Technology Review. Physica Medica **83**, 242–256 (2021)
4. Bodzas, A., Kodytek, P., Zidek, J.: A high-resolution large-scale dataset of pathological and normal white blood cells. Scientific Data **10**(1), 466 (2023)
5. Chen, C., Liao, T., Deng, X., Wu, Z., Huang, S., Zheng, Z.: Advances in Robust Federated Learning: A Survey with Heterogeneity Considerations. IEEE Transactions on Big Data pp. 1–20 (2025)
6. Collins, L., Hassani, H., Mokhtari, A., Shakkottai, S.: Exploiting Shared Representations for Personalized Federated Learning. In: Proceedings of International Conference on Machine Learning. pp. 2089–2099 (2021)
7. Ding, K., Feng, X., Yu, H.: To be Global or Personalized: Generalized Federated Learning with Cooperative Adaptation for Data Heterogeneity. Knowledge-Based Systems **301**, 112317 (2024)
8. Fan, Z., Yao, J., Han, B., Zhang, Y., Wang, Y., et al.: Federated Learning with Bilateral Curation for Partially Class-Disjoint Data. Advances in Neural Information Processing Systems **36** (2024)
9. Hsu, T.M.H., Qi, H., Brown, M.: Measuring the Effects of Non-Identical Data Distribution for Federated Visual Classification. arXiv preprint arXiv:1909.06335 (2019)

10. Jiang, Y., Wang, S., Valls, V., Ko, B.J., Lee, W.H., Leung, K.K., Tassiulas, L.: Model Pruning Enables Efficient Federated Learning on Edge Devices. IEEE Transactions on Neural Networks and Learning Systems **34**(12), 10374–10386 (2022)
11. Li, Q., He, B., Song, D.: Model-Contrastive Federated Learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10713–10722 (2021)
12. Li, T., Sahu, A.K., Zaheer, M., Sanjabi, M., Talwalkar, A., Smith, V.: Federated Optimization in Heterogeneous Networks. In: Proceedings of Machine Learning and Systems. vol. 2, pp. 429–450 (2020)
13. Li, X.C., Zhan, D.C.: FedRS: Federated Learning with Restricted Softmax for Label Distribution Non-IID Data. In: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. pp. 995–1005 (2021)
14. Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J.: A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems **33**(12), 6999–7019 (2022)
15. Liu, W., Zheng, Y., Xiang, Z., Wang, Y., Tian, Z., She, W.: An Efficient Federated Learning Method based on Enhanced Classification-GAN for Medical Image Classification. Multimedia Systems **31**(1), 1–17 (2025)
16. Luo, B., Li, X., Wang, S., Huang, J., Tassiulas, L.: Cost-Effective Federated Learning in Mobile Edge Networks. IEEE Journal on Selected Areas in Communications **39**(12), 3606–3621 (2021)
17. Matek, C., Schwarz, S., Spiekermann, K., Marr, C.: Human-level recognition of blast cells in acute myeloid leukaemia with convolutional neural networks. Nature Machine Intelligence **1**(11), 538–544 (2019)
18. Pei, J., Liu, W., Li, J., Wang, L., Liu, C.: A Review of Federated Learning Methods in Heterogeneous Scenarios. IEEE Transactions on Consumer Electronics (2024)
19. Reddi, S.J., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Konečný, J., Kumar, S., McMahan, H.B.: Adaptive Federated Optimization. In: Proceedings of International Conference on Learning Representations (2021)
20. Sabah, F., Chen, Y., Yang, Z., Raheem, A., Azam, M., Ahmad, N., Sarwar, R.: FairDPFL-SCS: Fair Dynamic Personalized Federated Learning with Strategic Client Selection for Improved Accuracy and Fairness. Information Fusion **115**, 102756 (2025)
21. Tan, Y., Long, G., Liu, L., Zhou, T., Lu, Q., Jiang, J., Zhang, C.: FedProto: Federated Prototype Learning across Heterogeneous Clients. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 8432–8440 (2022)
22. Tschandl, P., Rosendahl, C., Kittler, H.: The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
23. Wang, H., Yurochkin, M., Sun, Y., Papailiopoulos, D., Khazaeni, Y.: Federated Learning with Matched Averaging. arXiv preprint arXiv:2002.06440 (2020)
24. Wang, J., Liu, Q., Liang, H., Joshi, G., Poor, H.V.: Tackling the Objective Inconsistency Problem in Heterogeneous Federated Optimization. Advances in Neural Information Processing Systems **33**, 7611–7623 (2020)
25. Wang, L., Xu, S., Wang, X., Zhu, Q.: Addressing Class Imbalance in Federated Learning. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 10165–10173 (2021)
26. Wang, Y., Xu, Y., Shi, Q., Chang, T.H.: Quantized Federated Learning under Transmission Delay and Outage Constraints. IEEE Journal on Selected Areas in Communications **40**(1), 323–341 (2021)

27. Yan, Y., Zhu, L., Li, Y., Xu, X., Goh, R.S.M., Liu, Y., Khan, S., Feng, C.M.: A New Perspective to Boost Performance Fairness For Medical Federated Learning. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 13–23 (2024)
28. Ye, R., Xu, M., Wang, J., Xu, C., Chen, S., Wang, Y.: FedDisco: Federated Learning with Discrepancy-Aware Collaboration. In: Proceedings of International Conference on Machine Learning. pp. 39879–39902 (2023)
29. Zeng, S., Guo, P., Wang, S., Wang, J., Zhou, Y., Qu, L.: Tackling Data Heterogeneity in Federated Learning via Loss Decomposition. In: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 707–717 (2024)
30. Zhang, J., Li, Z., Li, B., Xu, J., Wu, S., Ding, S., Wu, C.: Federated Learning with Label Distribution Skew via Logits Calibration. In: Proceedings of International Conference on Machine Learning. pp. 26311–26329 (2022)
31. Zhu, Z., Shi, Y., Fan, P., Peng, C., Letaief, K.B.: ISFL: Federated Learning for Non-iid Data with Local Importance Sampling. IEEE Internet of Things Journal **11**(16), 27448–27462 (2024)
32. Zhu, Z., Hong, J., Zhou, J.: Data-Free Knowledge Distillation for Heterogeneous Federated Learning. In: Proceedings of International Conference on Machine Learning. pp. 12878–12889 (2021)