Journal of Machine Learning Research 26 (2025) 1-10 Submitted 12/23; Published 2/25

PFLlib: A Beginner-Friendly and Comprehensive Personalized Federated Learning Library and Benchmark

Jianqing Zhang[1] tsingz@sjtu.edu.cn
Yang Liu[2,*] liuy03@air.tsinghua.edu.cn
Yang Hua[3] y.hua@qub.ac.uk
Hao Wang[4] hwang9@stevens.edu
Tao Song[1] songt333@sjtu.edu.cn
Zhengui Xue[1] zhenguixue@sjtu.edu.cn
Ruhui Ma[1] ruhuima@sjtu.edu.cn
Jian Cao[1,*] cao-jian@sjtu.edu.cn
[1] *Shanghai Jiao Tong University, Shanghai, China*
[2] *Institute for AI Industry Research, Tsinghua University, Beijing, China*
[3] *Queen's University Belfast, Belfast, UK*
[4] *Stevens Institute of Technology, New Jersey, USA*
[*] *Corresponding Authors*

Editor: Albert Bifet

Abstract

Amid the ongoing advancements in Federated Learning (FL), a machine learning paradigm that allows collaborative learning with data privacy protection, personalized FL (pFL) has gained significant prominence as a research direction within the FL domain. Whereas traditional FL (tFL) focuses on jointly learning a global model, pFL aims to balance each client's global and personalized goals in FL settings. To foster the pFL research community, we started and built `PFLlib`, a comprehensive pFL library with an integrated benchmark platform. In `PFLlib`, we implemented 37 state-of-the-art FL algorithms (8 tFL algorithms and 29 pFL algorithms) and provided various evaluation environments with three statistically heterogeneous scenarios and 24 datasets. At present, `PFLlib`[1] has gained more than 1600 stars and 300 forks on GitHub.

Keywords: federated learning, personalization, privacy, benchmark, heterogeneity

1. Introduction

Federated Learning (FL) has gained significant attention due to its ability to perform distributed machine learning while ensuring privacy preservation (Yang et al., 2019). In traditional FL (tFL) algorithms, such as FedAvg (McMahan et al., 2017), participating clients train local models using local data and send only local model updates to a global server, which then aggregates these updates to obtain a global model. These approaches do not consider the customization needs of each local client. Personalized FL (pFL) is introduced to train customized client models to improve their performance on individualized tasks. In tandem with the burgeoning prominence of pFL, there has been a surge in the develop-

1. `https://www.pfllib.com/` and `https://github.com/TsingZ0/PFLlib`

ment of various pFL algorithms and associated techniques (Tan et al., 2022a; Zhang et al., 2023b). However, due to their rapid progress and diverse settings, the difficulties of tracking, implementing, and benchmarking these methods also grow tremendously.

To alleviate these challenges, we have developed `PFLlib`, a comprehensive pFL library with an integrated benchmark platform. `PFLlib` includes implementations of **37** state-of-the-art (SOTA) tFL and pFL algorithms, encompassing 8 tFL algorithms and **29** pFL algorithms. Our library is beginner-friendly and easily extensible, allowing contributors to seamlessly add new algorithms, scenarios, and datasets, thus ensuring that `PFLlib` remains up-to-date and popular. In addition, we have implemented three types of data heterogeneity scenarios and incorporated 24 datasets, covering Computer Vision (CV), Natural Language Processing (NLP), and Sensor Signal Processing (SSP) tasks. We can evaluate FL algorithms in `PFLlib` and assess their adaptability to various scenarios, providing valuable information for algorithm selection and evaluation in practical applications.

2. Related Work

With the rapid development of the FL field, numerous benchmarks and platforms have emerged in recent years. Most of their latest versions are for practical deployments, such as FATE (Liu et al., 2021), FedML (He et al., 2020), FederatedScope (Xie et al., 2023), Flower (Beutel et al., 2020), TensorFlow Federated[2], NVIDIA Clara[3], SecretFlow[4], Fedlearner[5], and PySyft[6]. Despite the efficient resource management and extensive functionality offered by these platforms, they can present a challenge for beginners who seek to comprehend the fundamental mechanisms of FL and delve into the philosophical aspects of existing FL algorithms. There are also some beginner-friendly platforms, such as LEAF (Caldas et al., 2018), NIID-Bench (Li et al., 2022), Motley (Wu et al., 2022b), OARF (Hu et al., 2022), FedEval (Chai et al., 2020), and FedLab (Zeng et al., 2023). However, these benchmarks and platforms still lack sufficient and up-to-date built-in SOTA pFL algorithms for researchers to learn, compare, and analyze.

pFL-Bench (Chen et al., 2022) is one of the latest PFL projects, but it supports only 5 SOTA pFL methods while the remaining pFL methods are variants created by combining them with existing approaches including FedBN (Li et al., 2021c), FedOpt (Asad et al., 2020), and Fine-tuning (FT). Besides, all the pFL algorithms in pFL-Bench are outdated (before 2022). In contrast, our `PFLlib` consists of 29 SOTA pFL algorithms. Moreover, due to our straightforward file structure, `PFLlib` is more accessible for beginners to learn and utilize pFL algorithms compared to the complex pFL-Bench.

3. `PFLlib`: A Beginner-Friendly and Comprehensive Library

Algorithms. In our `PFLlib`, the primary focus is on pFL algorithms. In addition, we have also included a selection of tFL algorithms to facilitate the evaluation of pFL algorithms, following previous pFL research (Li et al., 2021b; T Dinh et al., 2020; Zhang et al., 2023d).

2. https://www.tensorflow.org/federated
3. https://developer.nvidia.com/industries/healthcare
4. https://github.com/secretflow/secretflow
5. https://github.com/bytedance/fedlearner
6. https://github.com/OpenMined/PySyft

Based on their foundational techniques, we have categorized 8 tFL algorithms and 29 pFL algorithms. The detailed classification is presented in Table 1.

Table 1: FL Algorithm Taxonomy in our `PFLlib`.

	Category	Algorithms
8 tFL Algorithms	Basic tFL	FedAvg (McMahan et al., 2017)
	Update-correction-based tFL	SCAFFOLD (Karimireddy et al., 2020)
	Regularization-based tFL	FedProx (Li et al., 2020) and FedDyn (Acar et al., 2021)
	Model-splitting-based tFL	MOON (Li et al., 2021a) and FedLC (Zhang et al., 2022)
	Knowledge-distillation-based tFL	FedGen (Zhu et al., 2021) and FedNTD (Lee et al., 2022)
29 pFL Algorithms	Meta-learning-based pFL	Per-FedAvg (Fallah et al., 2020)
	Regularization-based pFL	pFedMe (T Dinh et al., 2020) and Ditto (Li et al., 2021b)
	Personalized-aggregation-based pFL	APFL (Deng et al., 2020), FedFomo (Zhang et al., 2020), FedAMP (Huang et al., 2021), FedPHP (Li et al., 2021d), APLE (Luo and Wu, 2022), and FedALA (Zhang et al., 2023d)
	Model-splitting-based pFL	FedPer (Arivazhagan et al., 2019), LG-FedAvg (Liang et al., 2020), FedRep (Collins et al., 2021), FedRoD (Chen and Chao, 2021), FedBABU (Oh et al., 2022), FedGC (Niu and Deng, 2022), FedCP (Zhang et al., 2023c), GPFL (Zhang et al., 2023b), FedGH (Yi et al., 2023), DBE (Zhang et al., 2023a), FedCAC (Wu et al., 2023), and PFL-DA (Shi and Kontar, 2023)
	Knowledge-distillation-based pFL	FedDistill (Seo et al., 2022), FML (Shen et al., 2020), FedKD (Wu et al., 2022a), FedProto (Tan et al., 2022b), FedPCL (Tan et al., 2022c), and FedPAC (Xu et al., 2022)
	Other pFL	FedMTL (Seo et al., 2022) and FedBN (Li et al., 2021c)

Scenarios and Datasets. In `PFLlib`, we first consider two types of scenarios for data heterogeneity: *label skew* and *feature shift*, where different client datasets differ in label categories and feature categories, respectively (Zhang et al., 2023b). Both CV and NLP classification tasks are considered in these two scenarios. Moreover, we also introduce a *real world* scenario with naturally collected datasets from distributed sensors (HAR (Anguita et al., 2012) and PAMAP2 (Reiss and Stricker, 2012) for SSP tasks), hospitals (Camelyon17 (Koh et al., 2021) for CV tasks), and camera traps (iWildCam (Koh et al., 2021) for CV tasks) to evaluate algorithms' performance in realistic scenarios.

Privacy Evaluation. We implement the popular Deep Leakage from Gradients (DLG) attack (Zhu et al., 2019) and the Peak Signal-to-Noise Ratio (PSNR) metric (Wu et al., 2022c) to evaluate the privacy-preserving abilities of existing tFL/pFL algorithms.

Easy to Use and Extend. In `PFLlib`, algorithms are implemented by **three** critical files: `serverX.py` for server creation, `clientX.py` for client creation, and `main.py` for hyperparameter configuration. Here, " `X` " represents some algorithm name. We can simply create a new algorithm `Y` by only adding specific features to `serverY.py` and `clientY.py` and while utilizing the core APIs in `serverbase.py` and `clientbase.py` . To create a scenario, users simply need **one** command line and run the evaluation with another **one** command line. We show an example of using FedALA on MNIST (LeCun et al., 1998):

```
# generate a practical non-iid and unbalanced scenario using MNIST
python generate_MNIST.py noniid - dir # in ./dataset
# evaluate the FedALA algorithm using a CNN with default hyperparameters
python main.py -data MNIST -m CNN -algo FedALA -gr 2000 -did 0 # in ./system
```

Impacts. Our `PFLlib` is active and popular in the pFL community, as shown by the increasing number of GitHub stars, forks, and active discussions. Due to its simplicity and extensibility, numerous new platforms and projects have been built upon it, such as the FL-bench[7], the HtFLlib[8], and the FL-IoT[9]. Besides, the experiments of several latest SOTA methods (Zhang et al., 2023d,c,b,a, 2024a,b,c) are also conducted using our `PFLlib`.

Benchmark. Due to limited space here, we only evaluate 20 algorithms in two *label skew* scenarios following the default settings[10] of GPFL (Zhang et al., 2023b). Please refer to our official website[11] for more documents, details, and results. In Table 2, we use the 4-layer CNN (McMahan et al., 2017) for CV tasks on Fashion-MNIST (FMNIST) (Xiao et al., 2017), Cifar100, and Tiny-ImageNet (Chrabaszcz et al., 2017) (TINY for short) datasets and use the fastText (Joulin et al., 2017) for NLP tasks on AG News (Zhang et al., 2015) dataset. We also use ResNet-18 (He et al., 2016) on Tiny-ImageNet and denote it TINY*.

Table 2: The test accuracy (%) on the CV and NLP tasks in *label skew* settings.

Settings	**Pathological *Label Skew* Setting**			**Practical *Label Skew* Setting**				
	FMNIST	Cifar100	TINY	FMNIST	Cifar100	TINY	TINY*	AG News
FedAvg	80.41±0.08	25.98±0.13	14.20±0.47	85.85±0.19	31.89±0.47	19.46±0.20	19.45±0.13	87.12±0.19
FedProx	78.08±0.15	25.94±0.16	13.85±0.25	85.63±0.57	31.99±0.41	19.37±0.22	19.27±0.23	87.21±0.13
FedGen	79.76±0.60	20.80±1.00	13.82±0.09	84.90±0.31	30.96±0.54	19.39±0.18	18.53±0.32	89.86±0.83
Per-FedAvg	99.18±0.08	56.80±0.26	28.06±0.40	95.10±0.10	44.28±0.33	25.07±0.07	21.81±0.54	87.08±0.26
pFedMe	99.35±0.14	58.20±0.14	27.71±0.40	97.25±0.17	47.34±0.46	26.93±0.19	33.44±0.33	87.08±0.18
Ditto	99.44±0.06	67.23±0.07	39.90±0.42	97.47±0.04	52.87±0.64	32.15±0.04	35.92±0.43	91.89±0.17
APFL	99.41±0.02	64.26±0.13	36.47±0.44	97.25±0.08	46.74±0.60	34.86±0.43	35.81±0.37	89.37±0.86
FedFomo	99.46±0.01	62.49±0.22	36.55±0.50	97.21±0.02	45.39±0.45	26.33±0.22	26.84±0.11	91.20±0.18
FedAMP	99.42±0.03	64.34±0.37	36.12±0.30	97.20±0.06	47.69±0.49	27.99±0.11	29.11±0.15	83.35±0.05
APPLE	99.30±0.01	65.80±0.08	36.22±0.40	97.06±0.07	53.22±0.20	35.04±0.47	39.93±0.52	84.10±0.18
FedALA	99.57±0.01	67.83±0.06	40.31±0.30	97.66±0.02	55.92±0.03	40.54±0.02	41.94±0.02	92.45±0.10
FedPer	99.47±0.03	63.53±0.21	39.80±0.39	97.44±0.06	49.63±0.54	33.84±0.34	38.45±0.85	91.85±0.24
FedRep	99.56±0.03	67.56±0.31	40.85±0.37	97.56±0.04	52.39±0.35	37.27±0.20	39.95±0.61	92.25±0.20
FedRoD	99.52±0.05	62.30±0.02	37.95±0.22	97.52±0.04	50.94±0.11	36.43±0.05	37.99±0.26	92.16±0.12
FedBABU	99.41±0.05	66.85±0.07	40.72±0.64	97.46±0.07	55.02±0.33	36.82±0.45	34.50±0.62	95.86±0.41
FedCP	99.66±0.04	71.80±0.16	44.52±0.22	**97.89±0.05**	59.56±0.08	**43.49±0.04**	**44.18±0.21**	92.89±0.10
GPFL	**99.85±0.08**	71.78±0.26	**44.58±0.06**	97.81±0.09	61.86±0.31	43.37±0.53	43.70±0.44	**97.97±0.14**
FedDBE	99.74±0.04	**73.38±0.18**	42.89±0.29	97.69±0.05	**64.39±0.27**	43.32±0.37	42.98±0.52	96.87±0.18
FedDistill	99.51±0.03	66.78±0.15	37.21±0.25	97.43±0.04	49.93±0.23	30.02±0.09	29.88±0.41	85.76±0.09
FedProto	99.49±0.04	69.18±0.03	36.78±0.07	97.40±0.02	52.70±0.33	31.21±0.16	26.38±0.40	96.34±0.58

4. Conclusion

To support the rapidly evolving pFL research community, we built `PFLlib`, a beginner-friendly library that includes 37 cutting-edge tFL / pFL algorithms. Besides, we also built a benchmark platform in `PFLlib` with comprehensive features, datasets, and scenarios.

7. `https://github.com/KarhouTam/FL-bench/tree/c11efc286dab4565245da34d7300d5bb07b87a0a`
8. `https://github.com/TsingZ0/HtFLlib`
9. `https://github.com/TsingZ0/FL-IoT`
10. Due to frequent updates, some default settings and codes for scenario creation may change in `PFLlib`.
11. `https://www.pfllib.com/`

Acknowledgments

This work was supported by the National Key R&D Program of China under Grant No.2022ZD0160504, the Program of Technology Innovation of the Science and Technology Commission of Shanghai Municipality (Granted No. 21511104700), the Interdisciplinary Program of Shanghai Jiao Tong University (project number YG2024QNB05), Tsinghua University(AIR)-Asiainfo Technologies (China) Inc. Joint Research Center, and the National Natural Science Foundation of China (62472284).

References

Durmus Alp Emre Acar, Yue Zhao, Ramon Matas, Matthew Mattina, Paul Whatmough, and Venkatesh Saligrama. Federated Learning Based on Dynamic Regularization. In *International Conference on Learning Representations (ICLR)*, 2021.

Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge L Reyes-Ortiz. Human activity recognition on smartphones using a multiclass hardware-friendly support vector machine. In *Ambient Assisted Living and Home Care: 4th International Workshop, IWAAL 2012, Vitoria-Gasteiz, Spain, December 3-5, 2012. Proceedings 4*, pages 216–223. Springer, 2012.

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated Learning with Personalization Layers. *arXiv preprint arXiv:1912.00818*, 2019.

Muhammad Asad, Ahmed Moustafa, and Takayuki Ito. Fedopt: Towards communication efficiency and privacy preservation in federated learning. *Applied Sciences*, 10(8):2864, 2020.

Daniel J Beutel, Taner Topal, Akhil Mathur, Xinchi Qiu, Javier Fernandez-Marques, Yan Gao, Lorenzo Sani, Kwing Hei Li, Titouan Parcollet, Pedro Porto Buarque de Gusmão, et al. Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*, 2020.

Sebastian Caldas, Sai Meher Karthik Duddu, Peter Wu, Tian Li, Jakub Konečnỳ, H Brendan McMahan, Virginia Smith, and Ameet Talwalkar. Leaf: A benchmark for federated settings. *arXiv preprint arXiv:1812.01097*, 2018.

Di Chai, Leye Wang, Liu Yang, Junxue Zhang, Kai Chen, and Qiang Yang. Fedeval: A holistic evaluation framework for federated learning. *arXiv preprint arXiv:2011.09655*, 2020.

Daoyuan Chen, Dawei Gao, Weirui Kuang, Yaliang Li, and Bolin Ding. pfl-bench: A comprehensive benchmark for personalized federated learning. *Advances in Neural Information Processing Systems*, 35:9344–9360, 2022.

Hong-You Chen and Wei-Lun Chao. On Bridging Generic and Personalized Federated Learning for Image Classification. In *International Conference on Learning Representations (ICLR)*, 2021.

Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. A Downsampled Variant of Imagenet as an Alternative to the Cifar Datasets. *arXiv preprint arXiv:1707.08819*, 2017.

Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting Shared Representations for Personalized Federated Learning. In *International Conference on Machine Learning (ICML)*, 2021.

Yuyang Deng, Mohammad Mahdi Kamani, and Mehrdad Mahdavi. Adaptive Personalized Federated Learning. *arXiv preprint arXiv:2003.13461*, 2020.

Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. Personalized Federated Learning with Theoretical Guarantees: A Model-Agnostic Meta-Learning Approach. In *International Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Chaoyang He, Songze Li, Jinhyun So, Xiao Zeng, Mi Zhang, Hongyi Wang, Xiaoyang Wang, Praneeth Vepakomma, Abhishek Singh, Hang Qiu, et al. Fedml: A research library and benchmark for federated machine learning. *arXiv preprint arXiv:2007.13518*, 2020.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Sixu Hu, Yuan Li, Xu Liu, Qinbin Li, Zhaomin Wu, and Bingsheng He. The oarf benchmark suite: Characterization and implications for federated learning systems. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 13(4):1–32, 2022.

Yutao Huang, Lingyang Chu, Zirui Zhou, Lanjun Wang, Jiangchuan Liu, Jian Pei, and Yong Zhang. Personalized Cross-Silo Federated Learning on Non-IID Data. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2021.

Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. Bag of Tricks for Efficient Text Classification. In *Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 2017.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic Controlled Averaging for Federated Learning. In *International Conference on Machine Learning (ICML)*, 2020.

Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In *International conference on machine learning*, 2021.

Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

Gihun Lee, Minchan Jeong, Yongjin Shin, Sangmin Bae, and Se-Young Yun. Preservation of the global knowledge by not-true distillation in federated learning. *Advances in Neural Information Processing Systems*, 35:38461–38474, 2022.

Qinbin Li, Bingsheng He, and Dawn Song. Model-Contrastive Federated Learning. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021a.

Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-iid data silos: An experimental study. In *2022 IEEE 38th international conference on data engineering (ICDE)*, pages 965–978. IEEE, 2022.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated Optimization in Heterogeneous Networks. In *Conference on Machine Learning and Systems (MLSys)*, 2020.

Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and Robust Federated Learning Through Personalization. In *International Conference on Machine Learning (ICML)*, 2021b.

Xiaoxiao Li, Meirui JIANG, Xiaofei Zhang, Michael Kamp, and Qi Dou. FedBN: Federated Learning on Non-IID Features via Local Batch Normalization. In *International Conference on Learning Representations (ICLR)*, 2021c.

Xin-Chun Li, De-Chuan Zhan, Yunfeng Shao, Bingshuai Li, and Shaoming Song. FedPHP: Federated Personalization with Inherited Private Models. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML)*, 2021d.

Paul Pu Liang, Terrance Liu, Liu Ziyin, Nicholas B Allen, Randy P Auerbach, David Brent, Ruslan Salakhutdinov, and Louis-Philippe Morency. Think locally, act globally: Federated learning with local and global representations. *arXiv preprint arXiv:2001.01523*, 2020.

Yang Liu, Tao Fan, Tianjian Chen, Qian Xu, and Qiang Yang. Fate: An industrial grade platform for collaborative learning with data protection. *The Journal of Machine Learning Research*, 22(1):10320–10325, 2021.

Jun Luo and Shandong Wu. Adapt to Adaptation: Learning Personalization for Cross-Silo Federated Learning. In *International Joint Conference on Artificial Intelligence (IJCAI)*, 2022.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.

Yifan Niu and Weihong Deng. Federated Learning for Face Recognition With Gradient Correction. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2022.

Jaehoon Oh, SangMook Kim, and Se-Young Yun. FedBABU: Toward Enhanced Representation for Federated Image Classification. In *International Conference on Learning Representations (ICLR)*, 2022.

Attila Reiss and Didier Stricker. Introducing a new benchmarked dataset for activity monitoring. In *2012 16th international symposium on wearable computers*, pages 108–109. IEEE, 2012.

Hyowoon Seo, Jihong Park, Seungeun Oh, Mehdi Bennis, and Seong-Lyun Kim. 16 federated knowledge distillation. *Machine Learning and Wireless Communications*, page 457, 2022.

Tao Shen, Jie Zhang, Xinkang Jia, Fengda Zhang, Gang Huang, Pan Zhou, Kun Kuang, Fei Wu, and Chao Wu. Federated mutual learning. *arXiv preprint arXiv:2006.16765*, 2020.

Naichen Shi and Raed Al Kontar. Personalized federated learning via domain adaptation with an application to distributed 3d printing. *Technometrics*, 65(3):328–339, 2023.

Canh T Dinh, Nguyen Tran, and Tuan Dung Nguyen. Personalized Federated Learning with Moreau Envelopes. In *International Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

Alysa Ziying Tan, Han Yu, Lizhen Cui, and Qiang Yang. Towards Personalized Federated Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2022a. Early Access.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. Fedproto: Federated Prototype Learning across Heterogeneous Clients. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2022b.

Yue Tan, Guodong Long, Jie Ma, Lu Liu, Tianyi Zhou, and Jing Jiang. Federated learning from pre-trained models: A contrastive learning approach. *arXiv preprint arXiv:2209.10083*, 2022c.

Chuhan Wu, Fangzhao Wu, Lingjuan Lyu, Yongfeng Huang, and Xing Xie. Communication-efficient federated learning via knowledge distillation. *Nature communications*, 13(1):2032, 2022a.

Shanshan Wu, Tian Li, Zachary Charles, Yu Xiao, Ken Liu, Zheng Xu, and Virginia Smith. Motley: Benchmarking heterogeneity and personalization in federated learning. In *Workshop on Federated Learning: Recent Advances and New Challenges (in Conjunction with NeurIPS 2022)*, 2022b.

Xinghao Wu, Xuefeng Liu, Jianwei Niu, Guogang Zhu, and Shaojie Tang. Bold but cautious: Unlocking the potential of personalized federated learning through cautiously aggressive collaboration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023.

Yuezhou Wu, Yan Kang, Jiahuan Luo, Yuanqin He, Lixin Fan, Rong Pan, and Qiang Yang. Fedcg: Leverage conditional gan for protecting privacy and maintaining competitive performance in federated learning. In *International Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2022c.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

Yuexiang Xie, Zhen Wang, Dawei Gao, Daoyuan Chen, Liuyi Yao, Weirui Kuang, Yaliang Li, Bolin Ding, and Jingren Zhou. Federatedscope: A flexible federated learning platform for heterogeneity. *Proceedings of the VLDB Endowment*, 16(5):1059–1072, 2023.

Jian Xu, Xinyi Tong, and Shao-Lun Huang. Personalized federated learning with feature alignment and classifier collaboration. In *The Eleventh International Conference on Learning Representations*, 2022.

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10 (2):1–19, 2019.

Liping Yi, Gang Wang, Xiaoguang Liu, Zhuan Shi, and Han Yu. Fedgh: Heterogeneous federated learning with generalized global header. In *Proceedings of the 31st ACM International Conference on Multimedia*, 2023.

Dun Zeng, Siqi Liang, Xiangjing Hu, Hui Wang, and Zenglin Xu. Fedlab: A flexible federated learning framework. *J. Mach. Learn. Res.*, 24:100–1, 2023.

Jianqing Zhang, Yang Hua, Jian Cao, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, and Haibing Guan. Eliminating domain bias for federated learning in representation space. *Advances in Neural Information Processing Systems*, 36, 2023a.

Jianqing Zhang, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, Jian Cao, and Haibing Guan. Gpfl: Simultaneously learning global and personalized feature information for personalized federated learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5041–5051, 2023b.

Jianqing Zhang, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, and Haibing Guan. Fedcp: Separating feature information for personalized federated learning via conditional policy. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023c.

Jianqing Zhang, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, and Haibing Guan. FedALA: Adaptive Local Aggregation for Personalized Federated Learning. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2023d.

Jianqing Zhang, Yang Liu, Yang Hua, and Jian Cao. Fedtgp: Trainable global prototypes with adaptive-margin-enhanced contrastive learning for data and model heterogeneity in federated learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024a.

Jianqing Zhang, Yang Liu, Yang Hua, and Jian Cao. An upload-efficient scheme for transferring knowledge from a server-side pre-trained generator to clients in heterogeneous federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024b.

Jianqing Zhang, Yang Liu, Yang Hua, Jian Cao, and Qiang Yang. Fedl2g: Learning to guide local training in heterogeneous federated learning. *arXiv preprint arXiv:2410.06490*, 2024c.

Jie Zhang, Zhiqi Li, Bo Li, Jianghe Xu, Shuang Wu, Shouhong Ding, and Chao Wu. Federated learning with label distribution skew via logits calibration. In *International Conference on Machine Learning*, 2022.

Michael Zhang, Karan Sapra, Sanja Fidler, Serena Yeung, and Jose M Alvarez. Personalized Federated Learning with First Order Model Optimization. In *International Conference on Learning Representations (ICLR)*, 2020.

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-Level Convolutional Networks for Text Classification. In *International Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2015.

Ligeng Zhu, Zhijian Liu, and Song Han. Deep Leakage from Gradients. In *International Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Zhuangdi Zhu, Junyuan Hong, and Jiayu Zhou. Data-Free Knowledge Distillation for Heterogeneous Federated Learning. In *International Conference on Machine Learning (ICML)*, 2021.