TOWARDS OUT-OF-FEDERATION GENERALIZATION IN FEDERATED LEARNING SUPPLEMENTARY

Anonymous authors
Paper under double-blind review

A SOURCE CODE

The source code for evaluation and visualization can be found in an anonymized repository. Please visit: https://anonymous.4open.science/r/TFL-8390 to access this resource.

B ADDITIONAL EXPERIMENTAL RESULTS

B.1 EVALUATING OOF-RESILIENCY ON OFFICEHOME

We conduct additional experiments on **OfficeHome** (Venkateswara et al., 2017) dataset. It contains $15,588$ images from four domains: art, clipart, product, and real world. The task is a 65-class classification problem. Like PACS's experimental setup, we evenly split each domain into 5 subsets, yielding 20 subsets, and treat each subset as a client. We followed the common *leave-one-domain-out* experiment, where 3 domains are used (15 clients) for training and 1 domain (5 clients) for testing. We use ResNet50 (He et al., 2016) as our model and train the model for 100 communication rounds. Each local client optimized the model using stochastic gradient descent (SGD) with a learning rate of 0.01, a momentum of 0.9, weight decay of $5e^{-4}$, and a batch size of 64. The model is evaluated using classification accuracy.

Table 1: Accuracy on the **OfficeHome** dataset. We conduct experiments using a *leave-one-domain-out* approach, meaning each domain serves as the evaluation domain in turn. Existing methods typically consider each domain as an individual client (Liu et al., 2021; Nguyen et al., 2022). However, in order to simulate a large-scale distributed setting, we took a different approach by further dividing each domain into 5 subsets and treating each subset as a separate client. This increased the total number of clients to 20. *Our method outperformed others across all experimental settings*, demonstrating superior results.

Models		Backbone	OfficeHome				
			A	C	P	R	Average
Centralized Methods	Mixup (Xu et al., 2020)	ResNet50	64.7	54.7	77.3	79.2	69.0
	CORAL (Sun & Saenko, 2016)	ResNet50	64.4	55.3	76.7	77.9	68.6
Federated Learning Methods	FedAvg	ResNet50	24.10	23.16	40.19	43.47	32.73
	FedProx	ResNet50	23.16	23.47	41.08	42.66	32.59
	DRFA	ResNet50	25.29	23.98	41.23	42.35	33.21
	FedSR	ResNet50	23.51	22.93	39.30	41.48	31.81
	TFL (Ours)	ResNet50	**26.37**	**24.47**	**43.96**	**44.74**	**34.89**

B.2 DATA PRE-PROCESSING ON eICU

We follow (Huang et al., 2019) to predict patient mortality using drug features. These features pertain to the medications administered to patients during the initial 48 hours of their ICU stay. We've extracted pertinent patient and corresponding drug feature data from two primary sources: the 'medication.csv' and 'patient.csv' files. Our final dataset is a table with the dimension of 19000×1411. Each row in this matrix symbolizes a unique patient, while each column corresponds to a distinct medication.

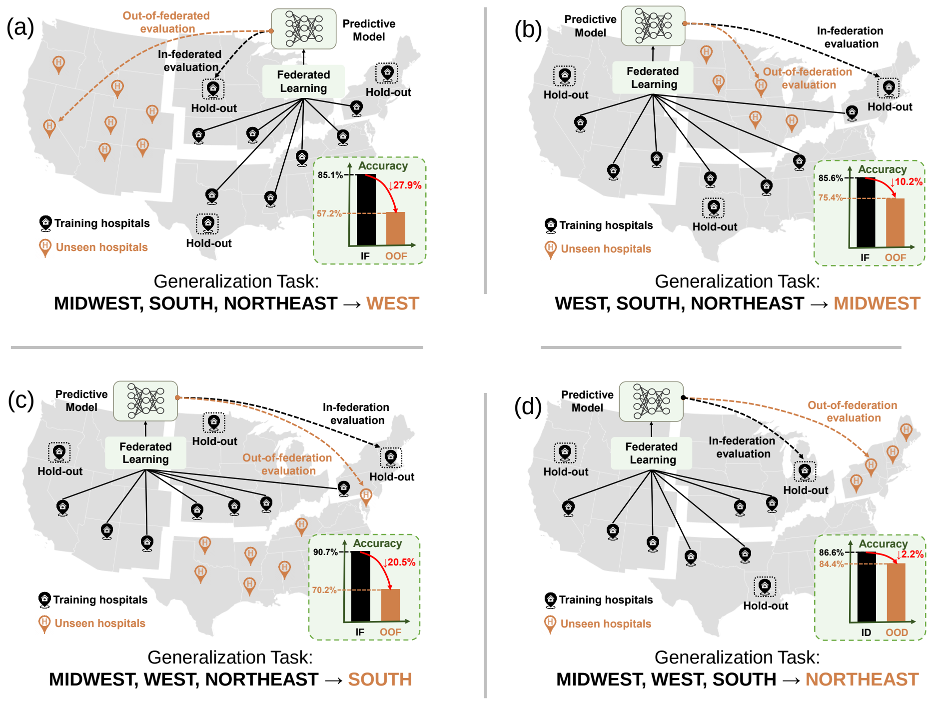

Figure 1: Additional results on eICU (ID *vs.* OOF performance). eICU's 72 hospitals are distributed across the United States. Specifically, there are 14 hospitals in the WEST, 28 hospitals in the MIDWEST, 26 hospitals in the SOUTH, and 4 hospitals in the NORTHEAST. We employ a *leave-one-region-out* approach, designating one geographic region as the OOF region while the remaining as ID regions. We observe a considerable gap between ID and OOF performance, indicating that *current FL methods are not robust against OOF data*.

B.3 More Results on eICU

We conduct more experiments on the eICU dataset to evaluate the gap between in-distribution (ID) and out-of-distribution (OOF) and visualize the results in Figure 1. *We observe that existing FL methods are not robust against OOF data.*

C Detailed Discussion of Related Work

Federated learning. Federated learning (Li et al., 2020a; Kairouz et al., 2021) has emerged as a powerful tool to protect data privacy in the distributed setting. It allows multiple clients/devices to collaborate in training a predictive model without sharing their local data. Despite the success, current FL methods are vulnerable to heterogeneous data (non-IID data) (Li et al., 2020b; Sattler et al., 2019), a common issue in real-world FL. Data heterogeneity posits significant challenges to FL, such as the severe convergence issue (Li et al., 2020b) and poor generalization ability to new clients (Sattler et al., 2019). To improve the model's robustness against data heterogeneity, `FedProx` add a proximal term to restrict the local model updating, avoiding biased models toward local data distribution. `SCAFFOLD` (Karimireddy et al., 2020) introduces a control variate to rectify the local update. `FedAlign` (Mendieta et al., 2022) improve the heterogeneous robustness by training local models with better generalization ability. However, most FL methods focus on the model's in-distribution performance. Orthogonal to existing work, we propose leveraging client relationships to improve the model's OOF generalization capability.

FL generalization to unseen clients. A handful of works tackle generalization to unseen clients in the FL setting. FedDG (Liu et al., 2021) is proposed to solve domain generalization in medical image classification. The key idea is to share the amplitude spectrum of images among local clients to augment the local data distributions. FedADG (Zhang et al., 2021) adopts the federated adversarial training to measure and align the local client distributions to a reference distribution.

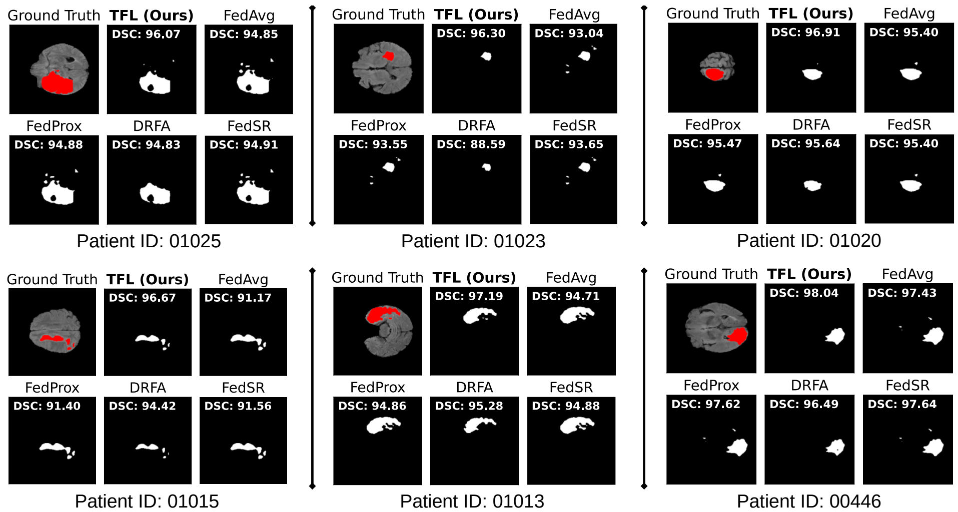

Figure 2: Additional qualitative results comparison on unseen patients of the FeTS dataset. We show both the tumor segmentation and DSC (↑) score. *Our method demonstrates consistent superior OOF-resiliency across a range of local demographics.*

FedGMA (Tenison et al., 2022) proposes gradient masking averaging to prioritize gradients aligned with the overall domain direction across clients. FedSR (Nguyen et al., 2022) proposes regularizing latent representation's ℓ_2 norm and class conditional information to enhance the OOF performance. However, existing methods often ignore scalability issues, yielding inferior performance in large-scale distributed setting (Bai et al., 2023). In this paper, we introduce an approach that employs client topology to strike a good balance between OOF-resiliency and scalability.

Graph topology learning. The problem of graph topology learning has been studied in different fields. In graph signal processing (Mateos et al., 2019; Dong et al., 2019; Stanković et al., 2020), existing work explore various way to learn the graph structure from data with structural regularization (*e.g.,* sparsity, smoothness, and community preservation (Zhu et al., 2021)). In Graph Neural Networks (GNNs) (Wu et al., 2020; Welling & Kipf, 2016), researchers have explored scenarios where the initial graph structure is unavailable, wherein a graph has to be estimated from objectives (Li et al., 2018; Norcliffe-Brown et al., 2018) or words (Chen et al., 2019; 2020). The existing graph topology learning methods often require centralizing the data, making it inapplicable in federated learning. However, how to estimate the graph topology with a privacy guarantee has been less investigated. In this paper, we explore simple methods to infer the graph topology using non-private information, *i.e.,* model weights.

D Additional Ablation Study

Hyperparameter q**.** We investigate the impact of hyperparameter η on eICU. Our findings demonstrate that setting $q = 0.1$ yields the best results. **Centrality.** We employed betweenness centrality to derive the topological prior. However, it is worth noting that other types of centrality, such as degree (Freeman et al., 2002) and closeness (Bavelas, 1950), could also be utilized. We conducted experiments on eICU to verify the impact of different centrality measures on TFL. Our findings indicate that betweenness centrality produces the best result. **Similarity metrics.** We investigate how the model performs under different similarity metrics on PACS. We found that the dot product-based metric produces the best results.

Table 2: Ablation study evaluating the efficacy of hyperparameter tuning, centrality., and similarity metric.

Effectiveness of q, ROC AUC ↑				
q =1.0	$q = 1e^{-1}$	$q = 1e^{-2}$	$q = 1e^{-3}$	$q = 1e^{-4}$
57.91	**58.31**	57.43	56.96	57.29
Effectiveness of centrality, ROC AUC ↑				
Betweenness	Degree	Closeness	Eigenvector	Current flow
58.28	57.69	57.86	57.57	57.83
Effectiveness of similarity measure, Accuracy ↑				
	ℓ_1	ℓ_2	dot produt	cosin
OOF Accuracy	58.11	58.26	59.14	58.52

Algorithm 1 Topology-aware Federated Learning

Input: K clients; learning rate η_θ and η_λ; communication round T; initial model $\theta^{(0)}$; initial $\lambda^{(0)}$; topology update frequency f .
while not convergence **do**
  for each communication round $t = 1, \cdots T$ **do**
    server **samples** m clients according to $\boldsymbol{\lambda}^{(t)}$
    for each client $i = 1, \cdots m$ in parallel **do**
      $\theta_i^{t+1} = \theta_i^t - \eta_{\theta_i^t} \nabla_{\theta_i^t} F(\theta_i^t)$
      client i send θ_i^{t+1} back to the server
    end for
    server **computes** $\boldsymbol{\theta}^{t+1} = \sum_{i=1}^m \theta_i^{t+1}$
    if $t\%f == 0$ **then**
      Updating graph $\mathcal{G}$ via Equation 4
    end if
    calculating topological prior p from $\mathcal{G}$
    calculating $\nabla_{\lambda^{(t)}} F(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\lambda}^{(t)})$ via Equation 6
    $\boldsymbol{\lambda}^{t+1} = \mathcal{P}_{\Delta_K}(\boldsymbol{\lambda}^t + \eta_\lambda^t \nabla_{\lambda^{(t)}} F(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\lambda}^{(t)}))$
  end for
end while

E Implementation Details

Experiment settings and evaluation metrics. For the **PACS dataset**, we evenly split each domain into 5 subsets, yielding 20 subsets, and we treat each subset as a client. We followed the common "leave-one-domain-out" experiment, where 3 domains are used (15 clients) for training and 1 domain (5 clients) for testing. We evaluated the model's performance using classification accuracy. We use ResNet18 (He et al., 2016) as our model and train the model for 100 communication rounds. Each local client optimized the model using stochastic gradient descent (SGD) with a learning rate of 0.01, momentum of 0.9, weight decay of $5e^{-4}$, and a batch size of 8. For **CIFAR-10/100**, we adopt the same model architecture as FedAvg (McMahan et al., 2017). The model has 2 convolution layers with 32, 64 5×5 kernels, and 2 fully connected layers with 512 hidden units. we use Dirichlet distribution (Hsu et al., 2019) to partition the dataset into the heterogeneous setting with 25 and 50 clients. For the **eICU dataset**, we treat each hospital as a client. We use a network of three fully connected layers. This architecture is similar to (Huang et al., 2019; Sheikhalishahi et al., 2020). We train our model for 30 communication rounds, using a batch size of 64 and a learning rate of 0.01, and report the performance on unseen hospitals. Within each communication round, clients performs 5 epochs (E = 5) of local optimization using SGD. The evaluation metric employed was the ROC-AUC, a common practice in eICU (Huang et al., 2019). For the **FeTS dataset**, we treat each institution as a client. We adopt the widely used U-Net (Ronneberger et al., 2015) model. We train our model for 20 communication rounds, using a learning rate of 0.01 and a batch size of 64. We conduct training with 16 intuitions and report results on 5 unseen institutions. Each institution performs 2 epochs of local optimization (E = 2) using SGD. The evaluation metric is Dice Similarity Coefficient (DSC $\uparrow$). For **TPT-48**, we consider two generalization tasks: (1) E(24) $\rightarrow$ W(24): Using the 24 eastern states as IF clients and the 24 western states as OOF clients; (2) N(24) $\rightarrow$ S(24): Using the 24 northern states as IF clients and the 24 southern states as OOF clients. We use a model similar to (Xu et al., 2022), which has 8 fully connected layers with 512 hidden units. We use SGD optimizer with a fixed momentum of 0.9. The evaluation metric is Mean Squared Error (MSE $\downarrow$). Algorithm 1 shows the overall algorithm of TFL. In implementation, we used dot product as the metric to measure client similarity.

F Discussion of Limitations

In this section, we discuss the limitations of TFL and the potential solutions.

Concerns on privacy leakage. Client topology learning may raise concerns about (unintentional) privacy leakage. However, we argue that any such leakage would be a general issue for FL methods rather than a unique concern for our approach. In comparison to standard FL, our method does not require additional information to construct the client topology, thus providing no worse privacy guarantees than well-established methods like FedAvg (McMahan et al., 2017) and FedProx (Li et al., 2020b). Nonetheless, FL may still be vulnerable to attacks that aim to extract sensitive information (Bhowmick et al., 2018; Melis et al., 2019). In future work, we plan to explore methods for mitigating (unintentional) privacy leakage.

Concerns on high-dimensional node embedding. As outlined in Section 3.1, we harness model weights as node embeddings. Nevertheless, incorporating large-scale models, such as Transformers (Vaswani et al., 2017; Dosovitskiy et al., 2021), may present a formidable obstacle, producing an overwhelmingly high-dimensional node vector. This will significantly increase computational demands for assessing node similarity. We argue that this can be addressed by dimension reduction. There are two possible ways: ① Utilizing model weights of certain layers as node embedding instead of the whole model. ② Directly learning the low-dimensional node embedding. One simple idea is to leverage Hypernetworks (Shamsian et al., 2021) to learn the node embedding with controllable dimensions.

Concerns on high computation cost for cross-device FL. Client topology learning requiris $\mathcal{O}(N^2)$ computation complexity for N clients. This quadratic complexity is prohibitively expensive in cross-device FL, where hundreds, thousands, or millions of clients/devices may be involved. In this case, we argue that the computation cost can be significantly reduced by client clustering (Sattler et al., 2020; Ghosh et al., 2020). By partitioning the clients into clusters, the total number of "clients" is reduced, allowing for cluster-level client topology learning to estimate the topology with reduced computation costs. We conducted experiments on the eICU dataset to empirically validate the effectiveness of our clustering-based method. The eICU dataset was selected for its large scale (72 clients) compared to all other evaluated datasets. Specifically, during client topology learning, we use KMeans (Lloyd, 1982) to partition the training clients into several (*e.g,* 10) clusters and learn the client topology at the cluster level. As shown in Table 3, our clustering approach significantly reduces computation costs by 69%, with only a small decrease in OOF performance by 0.77%.

Table 3: Compassion of computation of wall-clock time on eICU dataset. *Our clustering approach significantly reduces computation costs by 69%, with only a small decrease in OOF performance by 0.77%.*

	ROC-AUC	Wall-clock time (s)
FedAvg	57.18 ± 0.03	120.15
TFL	58.41 ± 0.06	437.61
TFL w/ Clustering	57.96 ± 0.18	133.08

REFERENCES

Ruqi Bai, Saurabh Bagchi, and David I. Inouye. Benchmarking algorithms for domain generalization in federated learning, 2023. URL https://openreview.net/forum?id=IsCg7qoy8i9.

Alex Bavelas. Communication patterns in task-oriented groups. *The journal of the acoustical society of America*, 22(6):725–730, 1950.

Abhishek Bhowmick, John Duchi, Julien Freudiger, Gaurav Kapoor, and Ryan Rogers. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018. URL https://arxiv.org/abs/1812.00984.

Yu Chen, Lingfei Wu, and Mohammed J Zaki. Graphflow: Exploiting conversation flow with graph neural networks for conversational machine comprehension. *arXiv preprint arXiv:1908.00059*, 2019.

Yu Chen, Lingfei Wu, and Mohammed J. Zaki. Reinforcement learning based graph-to-sequence model for natural question generation. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=HygnDhEtvr.

Xiaowen Dong, Dorina Thanou, Michael Rabbat, and Pascal Frossard. Learning graphs from data: A signal representation perspective. *IEEE Signal Processing Magazine*, 36(3):44–63, 2019. URL https://ieeexplore.ieee.org/document/8700665.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Linton C Freeman et al. Centrality in social networks: Conceptual clarification. *Social network: critical concepts in sociology. Londres: Routledge*, 1:238–263, 2002.

Avishek Ghosh, Jichan Chung, Dong Yin, and Kannan Ramchandran. An efficient framework for clustered federated learning. *Advances in Neural Information Processing Systems*, 33:19586–19597, 2020. URL https://arxiv.org/abs/2006.04088.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Tzu-Ming Harry Hsu, Hang Qi, and Matthew Brown. Measuring the effects of non-identical data distribution for federated visual classification. *arXiv preprint arXiv:1909.06335*, 2019.

Li Huang, Andrew L Shea, Huining Qian, Aditya Masurkar, Hao Deng, and Dianbo Liu. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *Journal of biomedical informatics*, 99:103291, 2019.

Peter Kairouz, H Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and open problems in federated learning. *Foundations and Trends® in Machine Learning*, 14(1–2):1–210, 2021. URL https://www.nowpublishers.com/article/Details/MAL-083.

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*, pp. 5132–5143. PMLR, 2020.

Ruoyu Li, Sheng Wang, Feiyun Zhu, and Junzhou Huang. Adaptive graph convolutional neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018. URL https://ojs.aaai.org/index.php/AAAI/article/view/11691.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated learning: Challenges, methods, and future directions. *IEEE Signal Processing Magazine*, 37(3):50–60, 2020a. URL https://ieeexplore.ieee.org/abstract/document/9084352.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *Proceedings of Machine Learning and Systems*, volume 2, pp. 429–450, 2020b. URL https://arxiv.org/abs/1812.06127.

Quande Liu, Cheng Chen, Jing Qin, Qi Dou, and Pheng-Ann Heng. Feddg: Federated domain generalization on medical image segmentation via episodic learning in continuous frequency space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1013–1023, 2021.

Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2): 129–137, 1982.

Gonzalo Mateos, Santiago Segarra, Antonio G Marques, and Alejandro Ribeiro. Connecting the dots: Identifying network structure via graph signal processing. *IEEE Signal Processing Magazine*, 36(3):16–43, 2019. URL https://ieeexplore.ieee.org/document/8700659?arnumber=8700659.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pp. 1273–1282. PMLR, 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *2019 IEEE symposium on security and privacy (SP)*, pp. 691–706. IEEE, 2019. URL https://ieeexplore.ieee.org/document/8835269.

Matias Mendieta, Taojiannan Yang, Pu Wang, Minwoo Lee, Zhengming Ding, and Chen Chen. Local learning matters: Rethinking data heterogeneity in federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8397–8406, 2022.

A Tuan Nguyen, Philip Torr, and Ser-Nam Lim. Fedsr: A simple and effective domain generalization method for federated learning. In *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/pdf?id=mrt90D00aQX.

Will Norcliffe-Brown, Stathis Vafeias, and Sarah Parisot. Learning conditioned graph structures for interpretable visual question answering. *Advances in neural information processing systems*, 31, 2018.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pp. 234–241. Springer, 2015.

Felix Sattler, Simon Wiedemann, Klaus-Robert Müller, and Wojciech Samek. Robust and communication-efficient federated learning from non-iid data. *IEEE transactions on neural networks and learning systems*, 31(9):3400–3413, 2019. URL https://ieeexplore.ieee.org/document/8889996.

Felix Sattler, Klaus-Robert Müller, and Wojciech Samek. Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints. *IEEE transactions on neural networks and learning systems*, 32(8):3710–3722, 2020. URL https://ieeexplore.ieee.org/document/9174890.

Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *International Conference on Machine Learning*, pp. 9489–9502. PMLR, 2021. URL http://proceedings.mlr.press/v139/shamsian21a/shamsian21a.pdf.

Seyedmostafa Sheikhalishahi, Vevake Balaraman, and Venet Osmani. Benchmarking machine learning models on multi-centre eicu critical care dataset. *Plos one*, 15(7):e0235424, 2020.

Ljubiša Stanković, Danilo Mandic, Miloš Daković, Miloš Brajović, Bruno Scalzo, Shengxi Li, Anthony G Constantinides, et al. Data analytics on graphs part iii: machine learning on graphs, from graph topology to applications. *Foundations and Trends® in Machine Learning*, 13(4):332–530, 2020.

Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, pp. 443–450. Springer, 2016.

Irene Tenison, Sai Aravind Sreeramadas, Vaikkunth Mugunthan, Edouard Oyallon, Eugene Belilovsky, and Irina Rish. Gradient masked averaging for federated learning. *arXiv preprint arXiv:2201.11986*, 2022.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. URL https://arxiv.org/abs/1706.03762.

Hemanth Venkateswara, Jose Eusebio, Shayok Chakraborty, and Sethuraman Panchanathan. Deep hashing network for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5018–5027, 2017.

Max Welling and Thomas N Kipf. Semi-supervised classification with graph convolutional networks. In *J. International Conference on Learning Representations (ICLR 2017)*, 2016.

Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and S Yu Philip. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

Minghao Xu, Jian Zhang, Bingbing Ni, Teng Li, Chengjie Wang, Qi Tian, and Wenjun Zhang. Adversarial domain adaptation with domain mixup. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 6502–6509, 2020.

Zihao Xu, Hao He, Guang-He Lee, Bernie Wang, and Hao Wang. Graph-relational domain adaptation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=kcwyXtt7yDJ.

Liling Zhang, Xinyu Lei, Yichun Shi, Hongyu Huang, and Chao Chen. Federated learning with domain generalization. *arXiv preprint arXiv:2111.10487*, 2021.

Yanqiao Zhu, Weizhi Xu, Jinghao Zhang, Yuanqi Du, Jieyu Zhang, Qiang Liu, Carl Yang, and Shu Wu. A survey on graph structure learning: Progress and opportunities. *arXiv e-prints*, pp. arXiv–2103, 2021. URL https://arxiv.org/pdf/2103.03036.pdf.