

Federated Learning and Class Imbalances - Breast Lesion Segmentation in DCE-MRI

Breast cancer is the most prevalent malignancy affecting women, necessitating early and precise lesion segmentation for accurate diagnosis and treatment. For its visualisation, dynamic contrast-enhanced MRI (DCE-MRI) stands out as one of the best options due to its prominent soft-tissue contrast, and recent advances in deep learning have shown great promise in automating lesion segmentation. However, adoption remains limited by strict privacy regulations and the heavy demands of centralised training. To address these limitations, this study investigates federated learning under real-world heterogeneity, using FedProx optimisation [1] to mitigate performance loss.

As a first step, a simple 3-layer U-Net, a more complex U-Net [2], and a ResNet-based transfer model were evaluated in a centralised framework with tailored data augmentation. The best model was then deployed in a federated learning setup using both FedAvg and FedProx algorithms. To simulate clinical variability, statistical heterogeneity was modelled via unbalanced data splits across 2, 6, and 10 clients, and system heterogeneity through varying client training loads (1 and 5 epochs per round).

Initial results showed that although a complex U-Net achieved the highest Dice score of 0.575, its longer training time led to selecting the first U-Net with augmentation as the federated learning baseline, balancing performance (Dice 0.538, IoU 0.368) and efficiency.

In the federated context, a two-client setup with equal data splits revealed that non-IID distributions caused disparate centralised performance (Dice 0.454 vs. 0.259); under FedAvg, the weaker client improved ~5% while the stronger declined ~15%, showing that even limited heterogeneity can impair training.

Building on these results, the Table on the right summarises the best and final Dice and IoU scores, along with total training times, across all simulated scenarios and number of clients.

Under statistical heterogeneity, the two-client setup showed that FedAvg led to increasing test loss

Problem	# Clients	Method	Dice Best	Dice Final	IoU Best	IoU Final	Time (h:m:s)
Statistical	2	FedAvg	0.508	0.360	0.340	0.220	06:12:53
		FedProx	0.532	0.480	0.362	0.316	-
	6	FedAvg	0.476	0.278	0.312	0.161	05:57:11
		FedProx	0.539	0.521	0.369	0.352	-
	10	FedAvg	0.322	0.121	0.192	0.064	03:49:44
		FedProx	0.528	0.528	0.359	0.359	-
System	2	FedAvg	0.334	0.191	0.200	0.106	07:53:35
		FedProx	0.519	0.519	0.350	0.350	-

as the smaller client produced divergent updates. FedProx stabilised training, improving the global Dice score by 12% and raising Client 2's score from 0.186 to 0.349. Increasing the number of clients showed that FedProx reliably maintained stability, matching the final global Dice scores of centralised training despite more clients.

In the system heterogeneity scenario, Client 2's higher number of local updates caused overfitting, reducing performance under FedAvg (Dice 0.191). FedProx mitigated performance drops, approaching the centralised global Dice (0.519) while allowing Client 1 to maintain and Client 2 to substantially improve their individual scores under heterogeneous training.

Training times were mainly driven by the largest client's data, which accounted for 50% in the 2- and 6-client setups. Consequently, adding more clients was meant to speed up training, but ended up increasing delays due to resource scheduling.

Ultimately, these results show FedProx's clear advantage over FedAvg in managing both statistical and system heterogeneity, improving convergence and generalisation in challenging federated settings. Thus, robustness and accuracy can be maintained through careful algorithm design rather than strict data uniformity - highlighting strong potential for assessing privacy restrictions in medical imaging segmentation.

[1] Li, T., et al. (2020). Federated optimization in heterogeneous networks.

[2] Zhao, X., et al. (2023). BreastDM: A DCE-MRI dataset for breast tumor image segmentation and classification. Computers in Biology and Medicine, 164:107255.