Supplementary Material for

FedIA: Federated Medical Image Segmentation with Heterogeneous Annotation Completeness

Table 1: Detailed implementations of the other methods. If correct in method, only correct pixel with value 0, the same as FedIA.

Methods	Parameters
FedAvg	the same as FedIA
ELR	$\beta = 0.99,\ \lambda = 1$
NR-Dice	$\gamma = 1.5$
ADELE	$\lambda = 1,\ \tau = 0.8,\ \rho = 0.8,\ r = 0.8/0.9$ (**MS**/**LUNG**)
FedCorr	$T_1 = 10,\ T_2 = 140,\ T_3 = 150$
RMD	$t_1 = 10,\ u = 0.8,\ \tau = 0.005$
FedNoRo	T (warm up round) = 10/20 (**MS**/**LUNG**)

Table 2: Ablation results about the correction threshold τ. We found that a smaller threshold is more appropriate when the annotation completeness is lower, which indicates that the network is not so confident about its prediction when the annotations are incomplete.

	MS			
τ	$m = 0$	$m = 1$	$m = 2$	$m = 3$
	4/6/8/10	3/5/7/9	2/4/6/8	1/3/5/7
0.5	74.68	73.55	**71.04**	**66.71**
0.6	74.04	73.51	68.06	65.86
0.7	74.66	73.13	69.43	64.66
0.8	**74.73**	**74.03**	69.22	56.53

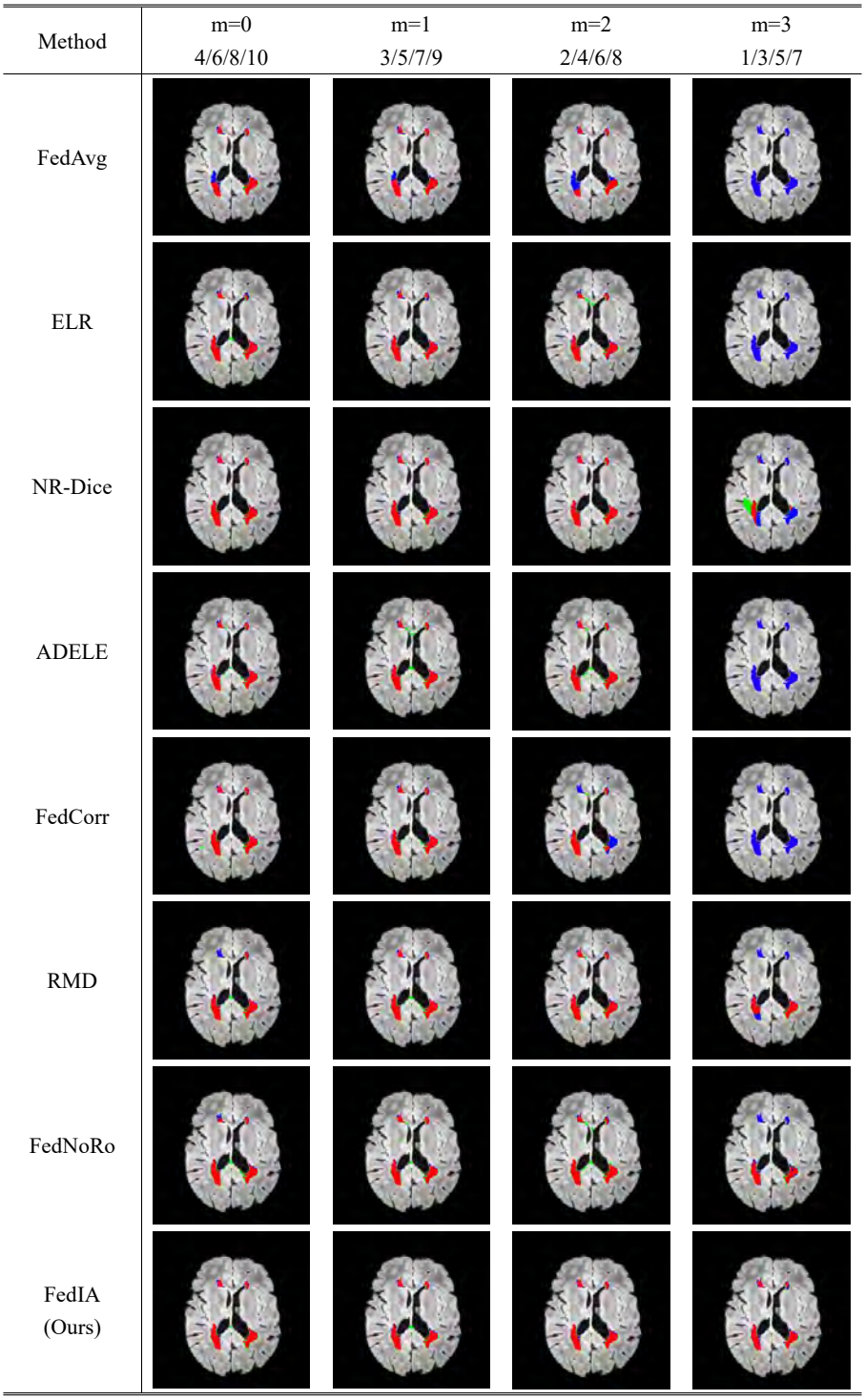

Fig. 1: Qualitative comparisons of different methods. We provide qualitative results in different settings on **MS** for visual comparison. Red, blue and green color show the prediction of true-positive, false-negative and false-positive regions, respectively.