Model	Training Method	Lesionwise Sensitivity	$p^{Default}$	p^{Method}	Lesionwise PPV	$p^{Default}$	p^{Method}
Default		0.25±0.04			0.93±0.03		
10	FT	0.55±0.02	<0.001*	0.769	0.72±0.02	<0.001*	<0.001*
	DN	0.54±0.02	<0.001*		0.28±0.02	<0.001*	
20	FT	0.59±0.02	<0.001*	<0.001*	0.73±0.02	<0.001*	<0.001*
	DN	0.50±0.01	<0.001*		0.45±0.03	<0.001*	
30	FT	0.59±0.01	<0.001*	<0.001*	0.75±0.02	<0.001*	<0.001*
	DN	0.50±0.01	<0.001*		0.52±0.03	<0.001*	
40	FT	0.60±0.02	<0.001*	<0.001*	0.71±0.02	<0.001*	<0.001*
	DN	0.54±0.02	<0.001*		0.59±0.02	<0.001*	
50	FT	0.63±0.02	<0.001*	0.549	0.71±0.02	<0.001*	0.008*
	DN	0.62±0.02	<0.001*		0.67±0.02	<0.001*	

Supplementary Table 2: Single timepoint FLAIR segmentation model performance with alternate deeper U-Net architecture. All sensitivity and PPV values are displayed as mean ± standard error. FT = fine-tuned, DN = de-novo, PPV = positive predictive value, $p^{Default}$ = p value from comparison with the default model, p^{Method} = p value from comparison between fine-tuned and de-novo training methods with the same training dataset size. * indicates statistical significance ($p < 0.05$ after accounting for Bonferroni correction).