	Average Accuracy			Worst-Group Accuracy		
	ERM	UW	DRO	ERM	UW	DRO
Waterbids	97.0 (0.2)	95.1 (0.3)	93.5 (0.3)	63.7 (1.9)	88.0 (1.3)	91.4 (1.1)
CelebA	94.9 (0.2)	92.9 (0.2)	92.9 (0.2)	47.8 (3.7)	83.3 (2.8)	88.9 (2.3)
MultiNLI	82.8 (0.1)	81.2 (0.1)	81.4 (0.1)	66.4 (1.6)	64.8 (1.6)	77.7 (1.4)