			Average Accuracy		Worst-Group Accuracy	
			ERM	DRO	ERM	DRO
Standard Regularization	Waterbirds	Train	100.0	100.0	100.0	100.0
		Test	97.3	97.4	60.0	76.9
	CelebA	Train	100.0	100.0	99.9	100.0
		Test	94.8	94.7	41.1	41.1
	MultiNLI	Train	99.9	99.3	99.9	99.0
		Test	82.5	82.0	65.7	66.4
Strong ℓ_2 Penalty	Waterbirds	Train	97.6	99.1	35.7	97.5
		Test	95.7	96.6	21.3	84.6
	CelebA	Train	95.7	95.0	40.4	93.4
		Test	95.8	93.5	37.8	86.7
Early Stopping	Waterbirds	Train	86.2	80.1	7.1	74.2
		Test	93.8	93.2	6.7	86.0
	CelebA	Train	91.3	87.5	14.2	85.1
		Test	94.6	91.8	25.0	88.3
	MultiNLI	Train	91.5	86.1	78.6	83.3
		Test	82.8	81.4	66.0	77.7