	Model	Multi-modality			Val. Acc.	Test Acc.
		Taxonomy	**Location**	**Time**		
	Empirical Risk Minimization (ERM) [30]				62.7 (2.4)	71.6 (2.5)
	CORAL [59]				60.3 (2.8)	73.3 (4.3)
	Group DRO [26]		–		60.0 (0.7)	72.7 (2.0)
	Fish [56]				58.0 (0.2)	63.2 (0.7)
	ABSGD [50]				–	72.7 (1.8)
	COSMO (no-context)		–		63.2 (0.4)	68.8 (2.1)
Single context	COSMO	✓			62.8 (2.2) (-0.4)	72.4 (2.5) (+3.6)
			✓		64.4 (1.0) (+1.2)	**74.5 (3.6)** (+5.7)
				✓	64.7 (0.4) (+1.5)	71.1 (3.1) (+2.3)
Multiple contexts	COSMO	✓	✓		**65.4 (0.4)** (+2.2)	70.4 (2.1) (+1.6)
		✓		✓	64.9 (1.6) (+1.7)	73.7 (3.8) (+4.9)
			✓	✓	63.0 (2.1) (-0.2)	74.2 (2.2) (+5.4)
		✓	✓	✓	65.0 (1.6) (+1.8)	71.5 (2.8) (+2.7)

Table 1. Species Classification results on iWildCam2020-WILDS (OOD) dataset. The first baseline in the second section shows the no-context baseline that uses only image-species labels as KG edges. All models use a pre-trained ResNet-50 as image encoder. Parentheses show standard deviation across 3 random seeds. Missing values are denoted by –.