Table 1: ID and OOD test grid accuracy with SEM (Standard Error of the Mean) as error bars for the *EnvGen* experiments part of settings G1-G5 and for the ResNet and Vanilla-ViT models.

Setting	Experiment	ResNet		Vanilla-ViT	
		ID	OOD	ID	OOD
G1	experiment-1	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0
	experiment-2	99.8 ± 0.0	96.5 ± 0.1	97.3 ± 0.2	55.4 ± 6.6
	experiment-3	99.8 ± 0.1	98.5 ± 0.2	97.5 ± 0.6	77.5 ± 3.4
	experiment-4	100.0 ± 0.0	99.9 ± 0.0	98.9 ± 0.6	77.9 ± 3.5
	experiment-5	99.9 ± 0.1	99.3 ± 0.1	98.2 ± 0.3	83.7 ± 1.4
G2	experiment-1	100.0 ± 0.0	99.9 ± 0.1	98.7 ± 0.3	1.9 ± 1.9
	experiment-2	88.9 ± 1.3	78.2 ± 7.2	57.1 ± 21.5	0.2 ± 0.2
	experiment-3	99.3 ± 0.3	98.2 ± 0.8	81.1 ± 11.3	8.2 ± 5.4
	experiment-4	100.0 ± 0.0	100.0 ± 0.0	96.3 ± 1.5	0.0 ± 0.0
	experiment-5	100.0 ± 0.0	100.0 ± 0.0	76.9 ± 2.3	0.2 ± 0.2
G3	experiment-1	100.0 ± 0.0	100.0 ± 0.0	100.0 ± 0.0	88.6 ± 9.6
	experiment-2	93.7 ± 0.5	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0
	experiment-3	96.2 ± 1.9	0.1 ± 0.1	2.3 ± 0.3	0.1 ± 0.0
	experiment-4	98.9 ± 1.0	0.0 ± 0.0	93.0 ± 2.2	0.0 ± 0.0
	experiment-5	99.4 ± 0.0	35.3 ± 2.7	92.6 ± 1.4	23.8 ± 3.3
G4	experiment-1	100.0 ± 0.0	23.7 ± 5.8	100.0 ± 0.0	90.6 ± 0.9
	experiment-2	30.7 ± 1.2	0.0 ± 0.0	3.8 ± 2.9	0.0 ± 0.0
	experiment-3	58.8 ± 2.9	0.1 ± 0.1	14.8 ± 1.0	0.0 ± 0.0
	experiment-4	95.2 ± 0.6	64.1 ± 4.0	51.9 ± 2.8	9.0 ± 1.4
	experiment-5	97.1 ± 0.9	36.9 ± 0.5	60.4 ± 14.7	9.4 ± 2.3
G5	experiment-1	100.0 ± 0.0	9.1 ± 2.0	97.7 ± 0.2	0.0 ± 0.0
	experiment-2	98.3 ± 0.3	0.0 ± 0.0	91.9 ± 3.3	0.0 ± 0.0
	experiment-3	98.3 ± 0.6	0.0 ± 0.0	32.7 ± 14.5	0.0 ± 0.0
	experiment-4	100.0 ± 0.0	0.0 ± 0.0	98.3 ± 0.9	0.0 ± 0.0
	experiment-5	99.7 ± 0.2	0.6 ± 0.2	41.9 ± 18.4	0.0 ± 0.0