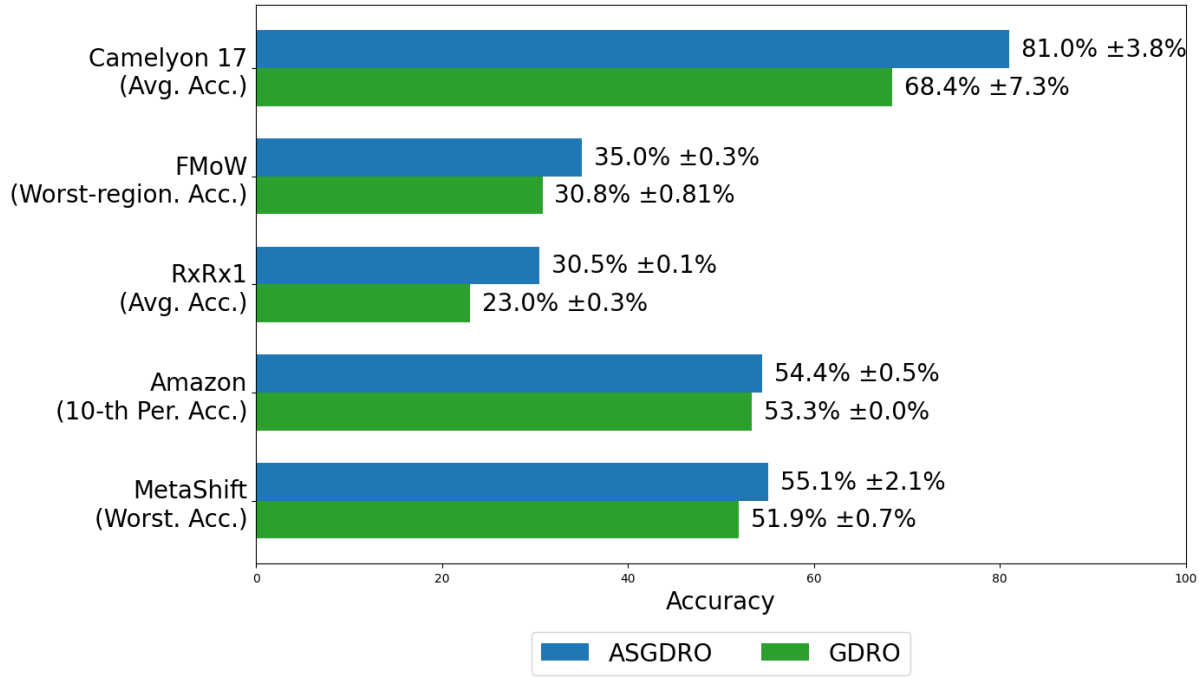

Figure 4: **GDRO and ASGDRO on Wilds Benchmark and MetaShift.** The parentheses below each name of dataset indicate the validation metric used for model selection[24]

Baseline	Method	PACS	VLCS	OfficeHome	TerraIncognita	DomainNet	Avg
Model Ensemble	ZooD-F.Selection† [12]	96.3	80.6	85.1	42.3	50.6	71.0
	EoA^{+}† [2]	93.2	80.4	80.2	55.2	**54.6**	72.7
	CADG^{+}†[9]	94.6	**82.2**	79.5	**55.7**	51.6	72.7
	Rew+HSIC†[6]	**96.7**	81.4	**85.3**	53.0	49.2	**73.1**
ResNet-50	ERM†	85.5	77.5	66.5	46.1	40.9	63.3
	IRM†	83.5	78.6	64.3	47.6	33.9	61.6
	GDRO†	84.4	76.7	66.0	43.2	33.3	60.7
	I-Mixup†	84.6	77.4	68.1	47.9	39.2	63.4
	MMD†	84.7	77.5	66.4	42.2	23.4	58.8
	SagNet†	86.3	77.8	68.1	48.6	40.3	64.2
	ARM†	85.1	77.6	64.8	45.5	35.5	61.7
	VREx†	84.9	78.3	66.4	46.4	33.6	61.9
	RSC†	85.2	77.1	65.5	46.6	38.9	62.7
	SWAD†[5]	**88.1**	**79.1**	**70.6**	**50.0**	**46.5**	**66.9**
CLIP-based	DPLCLIP	96.6	79.0	82.7	45.4	59.1	72.6
	DPLCLIP+GDRO	95.9	79.7	83.6	46.0	59.1	72.9
	DPLCLIP+ASGDRO	**96.8**	**80.7**	**83.7**	**48.9**	**59.8**	**74.0**

Table 4: **DomainBed.** The symbol † indicates reported performance. Except for the cited methods, all performances were taken from [17]. Bold accuracy w.r.t. average accuracy.

[17] Ishaan Gulrajani and David Lopez-Paz. In search of lost domain generalization. arXiv preprint arXiv:2007.01434, 2020.
[24] Pang Wei Koh, Shiori Sagawa, Henrik Marklund, Sang Michael Xie, Marvin Zhang, Akshay Balsubramani, Weihua Hu, Michihiro Yasunaga, Richard Lanas Phillips, Irena Gao, et al. Wilds: A benchmark of in-the-wild distribution shifts. In International Conference on Machine Learning, pages 5637–5664. PMLR,2021.