1 Exploring Different γ Values

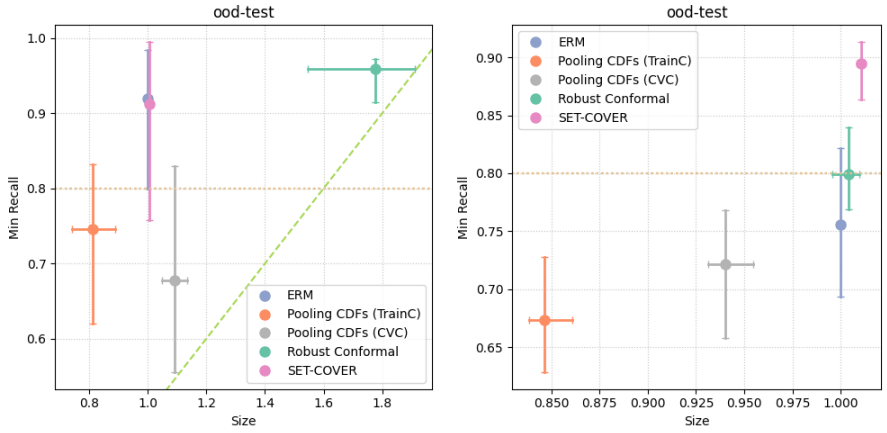
The γ parameter sets the desired recall level, and is supposed to be set in practice by practitioners according to task requirement. In our main experiments, which are presented at the body of this work, we targeted at a 0.9 recall level, which is associated with $\gamma = 0.1$. In this subsection we present results also for targeted recall levels of 0.8 and 0.95.

1.1 0.8 Recall

Table 1: Summary of OOD Results for recall level of 0.8 ($\gamma = 0.2$)

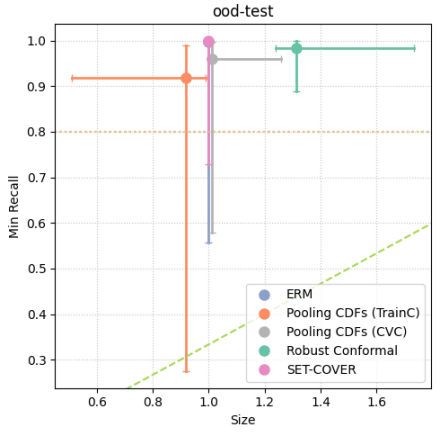
Model	**Camelyon**			**FMoW**		
	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.91	1.0	0.75	0.75	1.0	0.42
CDF Pooling-(TrainC)	0.74	0.81	0.38	0.67	0.84	0.05
CDF Pooling-(CVC)	0.67	1.09	0.45	0.72	0.94	0.16
Robust Conformal	0.95	1.77	0.90	0.79	1.00	0.50
SET-COVER	0.91	1.00	0.71	0.89	1.01	0.94

Model	**iWildCam**		
	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.99	1.0	0.71
CDF Pooling-(TrainC)	0.91	0.91	0.60
CDF Pooling-(CVC)	0.95	1.01	0.70
Robust Conformal	0.98	1.31	0.76
SET-COVER	0.99	1.00	0.71



(a) Camelyon

(b) FMoW



(c) iWildCam

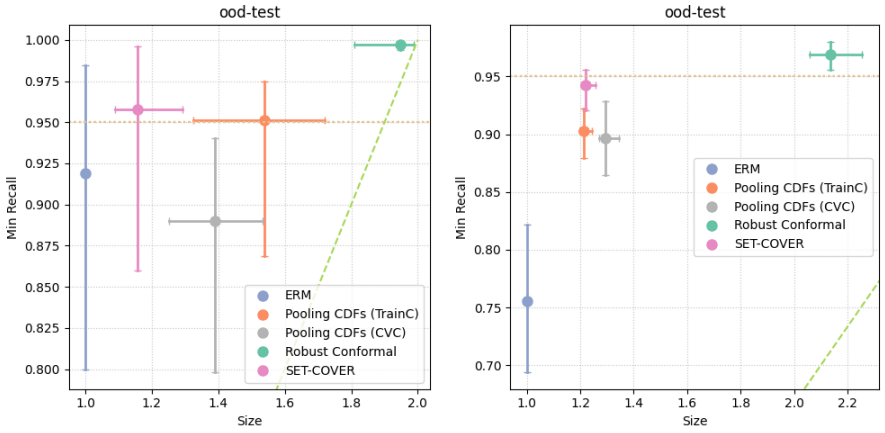
Figure 1: Results for recall target of 0.8 ($\gamma = 0.2$). Each figure represents Min-Recall over Avg Set Size cross. y-axis represents min-recall, and x-axis represents average set size. Each cross shows the median and the 25th and 75th percentiles for both metrics across domain. **Blue** represents ERM predictor, **Orange** represents Pooling CDFs (TrainC), **Grey** represents Pooling CDFs (CVC), **Green** represents robust conformal predictor, and **Pink** represents SET-COVER. The horizontal solid line represents the 90% recall target value, and dashed yellow diagonal line represents performance of a random predictor.

1.2 0.95 Recall

Table 2: Summary of OOD Results for recall level of 0.95 ($\gamma = 0.2$)

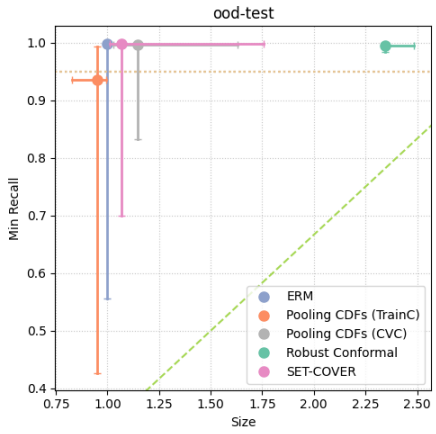
Model	Camelyon			FMoW		
	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.91	1.0	0.48	0.75	1.0	0.03
CDF Pooling-(TrainC)	0.95	1.53	0.5	0.90	1.21	0.07
CDF Pooling-(CVC)	0.88	1.39	0.26	0.89	1.29	0.14
Robust Conformal	0.99	1.94	0.93	0.96	2.13	0.64
SET-COVER	0.95	1.15	0.65	0.94	1.22	0.53

Model	iWildCam		
	Median Min Recall ↑	Median Avg Size ↓	Recall ≥ 90% Pctg ↑
ERM	0.99	1.0	0.70
CDF Pooling-(TrainC)	0.93	0.94	0.45
CDF Pooling-(CVC)	0.99	1.14	0.72
Robust Conformal	0.99	2.34	0.85
SET-COVER	0.99	1.07	0.77



(a) Camelyon

(b) FMoW



(c) iWildCam

Figure 2: Results for recall target of 0.95 ($\gamma = 0.05$). Each figure represents Min-Recall over Avg Set Size cross. y-axis represents min-recall, and x-axis represents average set size. Each cross shows the median and the 25th and 75th percentiles for both metrics across domain. **Blue** represents ERM predictor, **Orange** represents Pooling CDFs (TrainC), **Grey** represents Pooling CDFs (CVC), **Green** represents robust conformal predictor, and **Pink** represents SET-COVER. The horizontal solid line represents the 90% recall target value, and dashed yellow diagonal line represents performance of a random predictor.