APPENDIX

In this appendix, we present the experimental results of our proposed model stability based stopping criterion for extended models including kernelized models and non-gradient-based models which are introduced in Section 7. We left these detailed experimental results for this appendix due to the page limitation (25 pages).

A EXPERIMENTS FOR KERNELIZED MODELS

In this section, we validate our proposed algorithm on the kernelized models, including kernel logistic regression (kernel LR) and kernel SVMs with radial basis function (RBF) kernel and polynomial (PN) kernel (Section 7.1). Extensive experiments are implemented on both the UCI and ImageNet data sets with uncertainty sampling (US) strategy and expected model change (EMC) strategy, respectively.

A.1 Experiments with Uncertainty Sampling Strategy

We first present the experimental results coupled with US strategy. The following two subsections are detailed results for kernel LR and kernel SVMs, respectively.

A.1.1 Experiments for Kernel LR. Table 1-Table 2 present the results of kernel LR on UCI data sets with RBF kernel and PN kernel, respectively. Table 3-Table 4 present the results of kernel LR for ImageNet data sets. As shown in the tables, our proposed method MS is still observed to outperform the baselines most of the time. It achieves the highest accuracy with fewer annotations on most cases which demonstrates the effectiveness of our stopping criterion.

Table 1. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-LR with RBF kernel taking US as the strategy. The thresholds of MS are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Biodeg	*Size*	52.0	446.0	843.0	Δ_{Size}	**-101.0**	391.0	397.0	264.0	378.0	-88.0
	Acc	74.72%	87.45%	81.04%	Δ_{Acc}	**1.7%**	6.32%	6.42%	2.08%	2.26%	2.08%
Clave	*Size*	430.0	1393.0	6880.0	Δ_{Size}	**-86.0**	1135.0	5487.0	464.0	3354.0	344.0
	Acc	85.77%	98.42%	90.8%	Δ_{Acc}	**0.2%**	0.28%	7.62%	0.29%	0.52%	0.47%
Ionosphere	*Size*	17.0	87.0	280.0	Δ_{Size}	**-10.0**	-3.0	186.0	193.0	193.0	193.0
	Acc	75.21%	92.96%	84.79%	Δ_{Acc}	**2.54%**	2.82%	8.45%	8.17%	8.17%	8.17%
Spambase	*Size*	230.0	1873.0	3680.0	Δ_{Size}	**-407.0**	1308.0	1807.0	1807.0	1807.0	1807.0
	Acc	83.2%	94.26%	93.22%	Δ_{Acc}	**0.28%**	0.81%	1.04%	1.04%	1.04%	1.04%
WDBC	*Size*	28.0	93.0	454.0	Δ_{Size}	**-19.0**	5.0	283.0	101.0	138.0	23.0
	Acc	94.35%	98.26%	97.74%	Δ_{Acc}	**0.43%**	0.87%	3.74%	0.7%	0.61%	3.3%
D-vs-P	*Size*	80.0	139.0	1286.0	Δ_{Size}	**16.0**	22.0	1032.0	44.0	457.0	-59.0
	Acc	97.38%	99.64%	99.36%	Δ_{Acc}	**0.14%**	0.17%	0.3%	0.22%	0.19%	2.27%
E-vs-F	*Size*	77.0	134.0	1234.0	Δ_{Size}	**27.0**	**27.0**	1082.0	51.0	360.0	-42.0
	Acc	91.2%	99.61%	99.03%	Δ_{Acc}	**0.06%**	0.13%	0.52%	0.19%	0.26%	2.14%
M-vs-N	*Size*	78.0	161.0	1259.0	Δ_{Size}	**24.0**	30.0	1006.0	149.0	640.0	-83.0
	Acc	95.95%	99.4%	98.42%	Δ_{Acc}	**0.25%**	0.47%	0.95%	0.51%	0.79%	3.45%
U-vs-V	*Size*	78.0	113.0	1260.0	Δ_{Size}	**23.0**	32.0	1135.0	32.0	263.0	-31.0
	Acc	96.68%	99.92%	97.88%	Δ_{Acc}	**0.02%**	**0.02%**	2.04%	**0.02%**	0.11%	1.58%

Table 2. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-LR with PN kernel taking US as the strategy. The thresholds of MS are 2.5, 2.0, 1.0, 2.5, 2.5, 2.0, 2.5, 2.0, 2.0 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Biodeg	*Size*	52.0	272.0	843.0	Δ_{Size}	**319.0**	484.0	571.0	561.0	571.0	429.0
	Acc	63.92%	89.39%	84.43%	Δ_{Acc}	**2.83%**	4.25%	4.95%	4.72%	4.95%	3.54%
Clave	*Size*	430.0	1247.0	6880.0	Δ_{Size}	839.0	**430.0**	4236.0	4386.0	5139.0	1226.0
	Acc	96.22%	97.91%	96.98%	Δ_{Acc}	**0.19%**	0.23%	0.64%	0.9%	0.96%	0.33%
Ionosphere	*Size*	17.0	95.0	280.0	Δ_{Size}	158.0	185.0	**128.0**	185.0	185.0	185.0
	Acc	73.24%	88.03%	84.51%	Δ_{Acc}	**2.11%**	3.52%	2.82%	3.52%	3.52%	3.52%
Spambase	*Size*	230.0	1360.0	3680.0	Δ_{Size}	1630.0	2129.0	**-789.0**	2320.0	2320.0	2320.0
	Acc	77.87%	88.04%	82.49%	Δ_{Acc}	4.06%	5.94%	**2.7%**	5.55%	5.55%	5.55%
WDBC	*Size*	28.0	72.0	454.0	Δ_{Size}	**18.0**	152.0	322.0	368.0	375.0	332.0
	Acc	86.96%	98.43%	94.26%	Δ_{Acc}	**3.83%**	4.0%	4.09%	4.35%	4.26%	4.43%
D-vs-P	*Size*	80.0	160.0	1286.0	Δ_{Size}	**133.0**	173.0	1018.0	1126.0	1126.0	1066.0
	Acc	95.0%	98.73%	96.18%	Δ_{Acc}	**0.68%**	0.75%	2.36%	2.55%	2.55%	4.69%
E-vs-F	*Size*	77.0	272.0	1234.0	Δ_{Size}	**-165.0**	465.0	962.0	540.0	803.0	45.0
	Acc	93.69%	97.9%	96.6%	Δ_{Acc}	**0.49%**	0.81%	1.29%	1.46%	1.46%	0.65%
M-vs-N	*Size*	78.0	228.0	1259.0	Δ_{Size}	**234.0**	360.0	1031.0	1022.0	1029.0	848.0
	Acc	93.99%	97.11%	95.89%	Δ_{Acc}	**0.59%**	**0.59%**	1.23%	1.74%	1.23%	1.03%
U-vs-V	*Size*	78.0	259.0	1260.0	Δ_{Size}	**128.0**	736.0	1001.0	896.0	996.0	-181.0
	Acc	96.85%	99.68%	97.9%	Δ_{Acc}	**0.32%**	1.47%	1.79%	2.31%	1.89%	2.84%

Table 3. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-LR with RBF kernel taking US as the strategy. The thresholds of MS are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Basenji	*Size*	184.0	2108.0	2951.0	Δ_{Size}	**-222.0**	-481.0	518.0	592.0	407.0	703.0
	Acc	55.48%	65.49%	64.41%	Δ_{Acc}	**0.68%**	2.71%	2.84%	1.22%	1.35%	1.62%
Chihuahua	*Size*	175.0	2275.0	2800.0	Δ_{Size}	**-945.0**	-525.0	105.0	-140.0	455.0	70.0
	Acc	47.29%	69.14%	67.43%	Δ_{Acc}	**0.43%**	2.0%	1.0%	1.14%	2.0%	0.57%
English setter	*Size*	242.0	1712.0	3881.0	Δ_{Size}	**-490.0**	-686.0	196.0	931.0	1617.0	1813.0
	Acc	59.94%	64.68%	63.44%	Δ_{Acc}	**1.44%**	2.27%	2.16%	2.37%	2.57%	1.96%
German police dog	*Size*	174.0	1994.0	2785.0	Δ_{Size}	**-595.0**	-385.0	-280.0	420.0	490.0	70.0
	Acc	49.64%	64.71%	63.27%	Δ_{Acc}	**0.14%**	0.72%	1.29%	1.15%	1.72%	3.59%
Standard poodle	*Size*	195.0	3120.0	3129.0	Δ_{Size}	**-1170.0**	-1326.0	-351.0	-117.0	-78.0	-1287.0
	Acc	45.08%	75.99%	74.2%	Δ_{Acc}	**1.4%**	2.04%	2.17%	2.17%	2.55%	2.55%
Vizsla	*Size*	233.0	3288.0	3734.0	Δ_{Size}	282.0	47.0	-1034.0	**94.0**	188.0	-846.0
	Acc	60.28%	70.24%	69.7%	Δ_{Acc}	**1.39%**	3.1%	2.25%	3.32%	2.89%	1.61%
Yorkshire terrier	*Size*	304.0	4574.0	4874.0	Δ_{Size}	**-305.0**	-1647.0	300.0	-61.0	183.0	-1098.0
	Acc	68.28%	75.49%	74.84%	Δ_{Acc}	**0.16%**	0.9%	0.66%	1.39%	2.13%	0.74%
Wild dog	*Size*	242.0	2594.0	3875.0	Δ_{Size}	960.0	**-144.0**	-720.0	384.0	480.0	864.0
	Acc	65.84%	72.86%	72.03%	Δ_{Acc}	**0.31%**	0.62%	1.65%	3.61%	0.52%	1.75%
Wolf	*Size*	561.0	5713.0	8986.0	Δ_{Size}	**896.0**	-1904.0	-3839.0	1568.0	1680.0	-1232.0
	Acc	72.33%	76.87%	74.6%	Δ_{Acc}	3.29%	1.47%	5.66%	**0.53%**	2.58%	1.65%
Fox	*Size*	478.0	7582.0	7654.0	Δ_{Size}	-576.0	-1536.0	72.0	**-192.0**	**-192.0**	-1248.0
	Acc	72.52%	77.69%	76.96%	Δ_{Acc}	0.37%	0.57%	0.73%	**0.05%**	**0.05%**	0.21%
Rabbit	*Size*	349.0	3709.0	5593.0	Δ_{Size}	980.0	-1190.0	1050.0	1750.0	1884.0	**-490.0**
	Acc	69.12%	76.34%	74.2%	Δ_{Acc}	**0.64%**	0.86%	2.29%	3.57%	2.14%	1.14%
Cat	*Size*	732.0	7788.0	11722.0	Δ_{Size}	**3731.0**	3934.0	3934.0	3934.0	3934.0	3934.0
	Acc	64.68%	69.78%	65.98%	Δ_{Acc}	**2.08%**	3.8%	3.8%	3.8%	3.8%	3.8%
Panda	*Size*	351.0	2871.0	5628.0	Δ_{Size}	**1540.0**	-2030.0	1820.0	1680.0	2240.0	2170.0
	Acc	55.61%	84.52%	81.04%	Δ_{Acc}	**0.14%**	2.13%	2.98%	0.92%	2.27%	2.77%
Elephant	*Size*	392.0	3710.0	6282.0	Δ_{Size}	**316.0**	-2844.0	2572.0	869.0	1738.0	-2607.0
	Acc	80.47%	85.75%	83.21%	Δ_{Acc}	**0.25%**	2.42%	2.54%	0.57%	1.91%	3.69%

Table 4. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-LR with PN kernel taking US as the strategy. The thresholds of MS are all 2.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Basenji	*Size*	184.0	1689.0	2951.0	Δ_{Size}	**259.0**	1262.0	-604.0	1262.0	1262.0	1262.0
	Acc	51.96%	62.02%	59.27%	Δ_{Acc}	**1.58%**	2.75%	6.0%	2.75%	2.75%	2.75%
Chihuahua	*Size*	175.0	1692.0	2800.0	Δ_{Size}	**560.0**	1108.0	-1027.0	1108.0	1108.0	1108.0
	Acc	55.29%	63.9%	63.1%	Δ_{Acc}	**0.48%**	0.81%	7.52%	0.81%	0.81%	0.81%
English setter	*Size*	242.0	2864.0	3881.0	Δ_{Size}	**573.0**	**573.0**	-1923.0	1018.0	1018.0	1018.0
	Acc	53.58%	61.89%	60.04%	Δ_{Acc}	**1.57%**	4.76%	5.79%	1.85%	1.85%	1.85%
German police dog	*Size*	174.0	1376.0	2785.0	Δ_{Size}	828.0	1409.0	**128.0**	1409.0	1409.0	1409.0
	Acc	55.76%	61.88%	59.4%	Δ_{Acc}	**1.24%**	2.49%	5.88%	2.49%	2.49%	2.49%
Standard poodle	*Size*	195.0	2291.0	3129.0	Δ_{Size}	**419.0**	838.0	-1446.0	838.0	838.0	838.0
	Acc	62.41%	70.29%	69.31%	Δ_{Acc}	**0.68%**	0.98%	0.81%	0.98%	0.98%	0.98%
Vizsla	*Size*	233.0	2442.0	3734.0	Δ_{Size}	**928.0**	1292.0	-1927.0	1292.0	1292.0	1292.0
	Acc	55.41%	64.88%	62.26%	Δ_{Acc}	**1.5%**	2.62%	5.84%	2.62%	2.62%	2.62%
Yorkshire terrier	*Size*	304.0	3212.0	4874.0	Δ_{Size}	**-1891.0**	1662.0	324.0	1662.0	1662.0	1662.0
	Acc	65.22%	70.57%	68.58%	Δ_{Acc}	**1.07%**	1.99%	7.62%	1.99%	1.99%	1.99%
Wild dog	*Size*	242.0	2450.0	3875.0	Δ_{Size}	**528.0**	1425.0	-1344.0	1425.0	1425.0	1425.0
	Acc	57.89%	71.1%	66.87%	Δ_{Acc}	**1.14%**	4.23%	1.34%	4.23%	4.23%	4.23%
Wolf	*Size*	561.0	5601.0	8986.0	Δ_{Size}	**-1904.0**	3385.0	3385.0	3385.0	3385.0	3385.0
	Acc	64.41%	70.6%	68.37%	Δ_{Acc}	**1.56%**	2.22%	2.22%	2.22%	2.22%	2.22%
Fox	*Size*	478.0	1918.0	7654.0	Δ_{Size}	3072.0	5736.0	**2592.0**	5736.0	5736.0	5736.0
	Acc	65.57%	70.74%	69.38%	Δ_{Acc}	**0.31%**	1.36%	0.73%	1.36%	1.36%	1.36%
Rabbit	*Size*	349.0	4164.0	5593.0	Δ_{Size}	**-1348.0**	1429.0	330.0	1429.0	1429.0	1429.0
	Acc	55.5%	69.1%	66.14%	Δ_{Acc}	**2.2%**	2.97%	6.67%	2.97%	2.97%	2.97%
Cat	*Size*	732.0	10471.0	11722.0	Δ_{Size}	-9200.0	**1251.0**	-5672.0	**1251.0**	**1251.0**	**1251.0**
	Acc	49.62%	65.78%	65.39%	Δ_{Acc}	4.25%	**0.39%**	0.76%	**0.39%**	**0.39%**	**0.39%**
Panda	*Size*	351.0	3396.0	5628.0	Δ_{Size}	**679.0**	2232.0	1869.0	2232.0	2232.0	2232.0
	Acc	75.96%	77.77%	76.17%	Δ_{Acc}	**1.03%**	1.6%	1.1%	1.6%	1.6%	1.6%
Elephant	*Size*	392.0	3434.0	6282.0	Δ_{Size}	970.0	2849.0	**536.0**	2849.0	2849.0	2849.0
	Acc	69.94%	81.3%	77.89%	Δ_{Acc}	**2.42%**	3.4%	3.56%	3.4%	3.4%	3.4%

A.1.2 Experiments for Kernel SVM. Table 5-Table 6 present the results of kernel SVM on UCI data sets with RBF kernel and PN kernel, respectively. Table 7-Table 8 are the results of kernel SVM on ImageNet data sets. From these tables, similar results can be observed and MS performs best in these seven stopping criterions.

Table 5. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-SVM with RBF kernel taking US as the strategy. The thresholds of MS are 2.0, 2.0, 2.0, 2.0, 2.5, 2.0, 2.0, 2.0, 2.5 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Biodeg	*Size*	52.0	250.0	843.0	Δ_{Size}	107.0	152.0	240.0	367.0	501.0	84.0	**-29.0**
	Acc	77.6%	88.36%	86.87%	Δ_{Acc}	**1.19%**	1.43%	1.75%	1.58%	1.6%	1.38%	2.68%
Clave	*Size*	430.0	2620.0	6880.0	Δ_{Size}	-1250.0	-1221.0	350.0	-1456.0	-1193.0	-2190.0	**-957.0**
	Acc	96.16%	98.33%	98.21%	Δ_{Acc}	0.13%	0.14%	0.16%	0.29%	0.14%	2.17%	**0.11%**
Ionosphere	*Size*	17.0	64.0	280.0	Δ_{Size}	161.0	40.0	99.0	61.0	136.0	27.0	**6.0**
	Acc	78.73%	95.92%	94.65%	Δ_{Acc}	**1.06%**	1.76%	1.34%	1.27%	1.34%	1.97%	8.24%
Spambase	*Size*	230.0	1423.0	3680.0	Δ_{Size}	**-199.0**	-330.0	1708.0	-250.0	880.0	-391.0	343.0
	Acc	87.36%	93.53%	93.16%	Δ_{Acc}	0.61%	0.43%	**0.42%**	0.47%	0.25%	0.53%	0.45%
WDBC	*Size*	28.0	64.0	454.0	Δ_{Size}	366.0	40.0	318.0	28.0	80.0	**23.0**	101.0
	Acc	94.49%	98.72%	98.2%	Δ_{Acc}	**0.58%**	0.81%	0.75%	0.99%	1.04%	0.87%	0.93%
D-vs-P	*Size*	80.0	148.0	1286.0	Δ_{Size}	26.0	**18.0**	629.0	32.0	98.0	-68.0	152.0
	Acc	96.54%	99.53%	99.35%	Δ_{Acc}	**0.11%**	0.28%	0.17%	0.19%	0.17%	3.0%	0.26%
E-vs-F	*Size*	77.0	184.0	1234.0	Δ_{Size}	32.0	24.0	540.0	**19.0**	86.0	512.0	94.0
	Acc	94.94%	99.34%	99.06%	Δ_{Acc}	**0.16%**	0.26%	0.29%	0.29%	0.32%	0.84%	0.33%
M-vs-N	*Size*	78.0	171.0	1259.0	Δ_{Size}	380.0	**48.0**	707.0	54.0	163.0	-93.0	109.0
	Acc	96.65%	98.99%	98.8%	Δ_{Acc}	**0.19%**	0.32%	0.25%	0.25%	0.25%	2.34%	0.28%
U-vs-V	*Size*	78.0	140.0	1260.0	Δ_{Size}	285.0	22.0	601.0	**10.0**	89.0	-62.0	68.0
	Acc	97.05%	99.72%	99.57%	Δ_{Acc}	0.57%	0.16%	0.14%	**0.09%**	0.14%	2.67%	0.11%

Table 6. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-SVM with PN kernel taking US as the strategy. The thresholds of MS are 2.0, 2.0, 2.0, 1.5, 2.0, 2.0, 2.0, 2.0, 2.0 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Biodeg	*Size*	52.0	339.0	843.0	Δ_{Size}	-108.0	-33.0	43.0	66.0	271.0	**-50.0**	-103.0
	Acc	77.78%	89.15%	87.69%	Δ_{Acc}	1.42%	1.51%	2.26%	1.65%	1.65%	**1.27%**	1.46%
Clave	*Size*	430.0	1238.0	6880.0	Δ_{Size}	**-17.0**	-172.0	1926.0	-52.0	559.0	-808.0	163.0
	Acc	94.19%	97.85%	96.33%	Δ_{Acc}	**0.23%**	0.31%	0.9%	0.29%	0.52%	3.66%	0.37%
Ionosphere	*Size*	17.0	102.0	280.0	Δ_{Size}	**-10.0**	5.0	45.0	107.0	151.0	64.0	-17.0
	Acc	72.58%	90.23%	84.23%	Δ_{Acc}	**3.47%**	4.41%	4.88%	5.63%	5.45%	4.88%	4.41%
Spambase	*Size*	230.0	2134.0	3680.0	Δ_{Size}	-1104.0	-1063.0	934.0	-920.0	**23.0**	-1251.0	-1086.0
	Acc	87.08%	93.45%	93.11%	Δ_{Acc}	0.8%	0.65%	**0.47%**	0.54%	0.66%	0.66%	0.67%
WDBC	*Size*	28.0	60.0	454.0	Δ_{Size}	**0.0**	23.0	302.0	45.0	122.0	50.0	48.0
	Acc	93.45%	98.09%	96.81%	Δ_{Acc}	**0.7%**	1.16%	1.22%	1.28%	0.99%	1.16%	1.28%
D-vs-P	*Size*	80.0	223.0	1286.0	Δ_{Size}	-47.0	-58.0	510.0	-54.0	70.0	-93.0	**1.0**
	Acc	96.27%	99.42%	99.25%	Δ_{Acc}	0.35%	0.35%	**0.08%**	0.35%	0.27%	1.51%	0.33%
E-vs-F	*Size*	77.0	160.0	1234.0	Δ_{Size}	18.0	**8.0**	333.0	33.0	102.0	1075.0	57.0
	Acc	96.6%	99.48%	99.19%	Δ_{Acc}	**0.19%**	0.26%	0.39%	0.23%	**0.19%**	0.29%	0.23%
M-vs-N	*Size*	78.0	184.0	1259.0	Δ_{Size}	24.0	**-5.0**	498.0	83.0	131.0	1075.0	70.0
	Acc	95.32%	99.68%	99.43%	Δ_{Acc}	**0.16%**	0.7%	0.19%	0.25%	0.19%	0.25%	0.54%
U-vs-V	*Size*	78.0	136.0	1260.0	Δ_{Size}	28.0	27.0	492.0	**25.0**	82.0	-58.0	113.0
	Acc	97.01%	99.75%	99.62%	Δ_{Acc}	**0.13%**	0.17%	0.17%	0.19%	0.17%	2.73%	0.15%

Table 7. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-SVM with RBF kernel taking US as the strategy. The thresholds of MS are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Basenji	*Size*	184.0	1897.0	2951.0	Δ_{Size}	902.0	844.0	**-453.0**	1054.0	1054.0	1054.0	-370.0
	Acc	55.87%	64.19%	63.33%	Δ_{Acc}	**0.72%**	0.89%	3.27%	0.87%	0.87%	0.87%	1.79%
Chihuahua	*Size*	175.0	1829.0	2800.0	Δ_{Size}	**166.0**	464.0	-936.0	971.0	971.0	963.0	-359.0
	Acc	58.61%	65.21%	63.96%	Δ_{Acc}	**0.93%**	1.29%	2.25%	1.25%	1.25%	1.29%	2.47%
English setter	*Size*	242.0	2496.0	3881.0	Δ_{Size}	**743.0**	1209.0	-1372.0	1385.0	1385.0	1383.0	-1421.0
	Acc	56.42%	65.5%	64.04%	Δ_{Acc}	**1.07%**	1.36%	4.43%	1.45%	1.45%	1.44%	3.98%
German police dog	*Size*	174.0	1959.0	2785.0	Δ_{Size}	**410.0**	689.0	-899.0	826.0	826.0	826.0	-1155.0
	Acc	56.89%	63.79%	63.07%	Δ_{Acc}	**0.63%**	0.95%	3.32%	0.72%	0.72%	0.72%	3.6%
Standard poodle	*Size*	195.0	2188.0	3129.0	Δ_{Size}	**0.0**	-47.0	-120.0	917.0	933.0	-94.0	-733.0
	Acc	65.7%	74.64%	73.98%	Δ_{Acc}	**0.6%**	0.75%	1.34%	0.66%	0.65%	0.92%	2.25%
Vizsla	*Size*	233.0	3175.0	3734.0	Δ_{Size}	**-165.0**	-66.0	-1222.0	559.0	559.0	348.0	-780.0
	Acc	59.83%	68.63%	68.07%	Δ_{Acc}	**0.41%**	0.66%	3.1%	0.56%	0.56%	0.64%	1.09%
Yorkshire terrier	*Size*	304.0	3385.0	4874.0	Δ_{Size}	-299.0	-403.0	-868.0	1366.0	1450.0	**-220.0**	-610.0
	Acc	68.61%	74.81%	74.52%	Δ_{Acc}	0.82%	0.83%	1.89%	**0.26%**	0.3%	0.66%	0.66%
Wild dog	*Size*	242.0	2623.0	3875.0	Δ_{Size}	**67.0**	-422.0	202.0	1252.0	1252.0	403.0	-797.0
	Acc	66.5%	73.79%	73.23%	Δ_{Acc}	0.52%	0.76%	0.76%	0.56%	0.56%	**0.33%**	0.99%
Wolf	*Size*	561.0	5567.0	8986.0	Δ_{Size}	**302.0**	-829.0	2677.0	3270.0	3394.0	-314.0	784.0
	Acc	72.0%	77.39%	77.1%	Δ_{Acc}	0.52%	0.54%	0.51%	**0.28%**	**0.28%**	0.39%	0.37%
Fox	*Size*	478.0	6027.0	7654.0	Δ_{Size}	-1622.0	-1853.0	**146.0**	1555.0	1620.0	-1229.0	-749.0
	Acc	71.85%	77.62%	77.4%	Δ_{Acc}	0.71%	0.76%	0.44%	0.22%	**0.21%**	0.62%	0.34%
Rabbit	*Size*	349.0	4138.0	5593.0	Δ_{Size}	**-56.0**	-830.0	-427.0	1455.0	1455.0	-485.0	-737.0
	Acc	69.26%	77.15%	76.82%	Δ_{Acc}	**0.28%**	0.52%	1.42%	0.33%	0.33%	0.42%	0.73%
Cat	*Size*	732.0	7964.0	11722.0	Δ_{Size}	2587.0	-735.0	**626.0**	3758.0	3758.0	2705.0	1264.0
	Acc	64.69%	69.09%	68.77%	Δ_{Acc}	**0.14%**	0.46%	0.79%	0.33%	0.33%	0.24%	0.29%
Panda	*Size*	351.0	2570.0	5628.0	Δ_{Size}	**49.0**	119.0	2255.0	2296.0	2709.0	-105.0	490.0
	Acc	77.7%	83.03%	82.57%	Δ_{Acc}	**0.26%**	0.36%	0.65%	0.48%	0.43%	0.43%	0.42%
Elephant	*Size*	392.0	3355.0	6282.0	Δ_{Size}	-474.0	-474.0	1128.0	1264.0	2094.0	-869.0	**-382.0**
	Acc	78.94%	84.32%	84.13%	Δ_{Acc}	0.57%	0.54%	0.48%	**0.1%**	0.19%	0.89%	0.2%

Table 8. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-SVM with PN kernel taking US as the strategy. The thresholds of MS are 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 0.5, 1.0, 1.0, 0.5, 1.0, 0.5, 0.5 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Basenji	*Size*	184.0	2182.0	2951.0	Δ_{Size}	**-229.0**	769.0	-1354.0	769.0	769.0	725.0	-1177.0
	Acc	55.51%	65.12%	63.9%	Δ_{Acc}	**1.0%**	1.22%	3.27%	1.22%	1.22%	1.65%	2.98%
Chihuahua	*Size*	175.0	1848.0	2800.0	Δ_{Size}	**-7.0**	875.0	-490.0	952.0	952.0	896.0	-840.0
	Acc	55.83%	66.11%	64.43%	Δ_{Acc}	**0.83%**	1.6%	1.74%	1.69%	1.69%	1.74%	4.71%
English setter	*Size*	242.0	2625.0	3881.0	Δ_{Size}	**121.0**	1256.0	-443.0	1256.0	1256.0	1231.0	-1187.0
	Acc	52.5%	64.32%	62.84%	Δ_{Acc}	**0.84%**	1.47%	2.08%	1.47%	1.47%	1.73%	2.29%
German police dog	*Size*	174.0	2141.0	2785.0	Δ_{Size}	**7.0**	220.0	-1130.0	644.0	644.0	636.0	-994.0
	Acc	52.8%	64.36%	62.61%	Δ_{Acc}	**1.69%**	3.53%	6.05%	1.75%	1.75%	2.09%	3.3%
Standard poodle	*Size*	195.0	1762.0	3129.0	Δ_{Size}	**-52.0**	169.0	187.0	1320.0	1362.0	137.0	-806.0
	Acc	62.01%	75.31%	74.24%	Δ_{Acc}	**0.4%**	1.06%	2.23%	1.0%	1.21%	0.7%	3.19%
Vizsla	*Size*	233.0	3344.0	3734.0	Δ_{Size}	-893.0	390.0	-715.0	380.0	390.0	**-75.0**	-1626.0
	Acc	60.45%	70.75%	69.81%	Δ_{Acc}	1.71%	0.94%	1.8%	1.13%	0.94%	**0.73%**	3.34%
Yorkshire terrier	*Size*	304.0	3064.0	4874.0	Δ_{Size}	**-442.0**	-336.0	104.0	1571.0	1781.0	31.0	-671.0
	Acc	69.24%	75.86%	74.67%	Δ_{Acc}	**0.55%**	0.74%	1.48%	0.72%	1.05%	1.17%	1.09%
Wild dog	*Size*	242.0	3346.0	3875.0	Δ_{Size}	**-1216.0**	-1216.0	-1269.0	464.0	529.0	-720.0	-1232.0
	Acc	64.71%	74.48%	73.82%	Δ_{Acc}	**0.48%**	0.76%	2.17%	0.79%	0.65%	0.86%	1.2%
Wolf	*Size*	561.0	6833.0	8986.0	Δ_{Size}	**-2296.0**	-2128.0	2153.0	1792.0	2128.0	-1624.0	-1624.0
	Acc	73.0%	77.36%	76.69%	Δ_{Acc}	**0.4%**	0.53%	0.67%	0.67%	0.73%	0.44%	0.67%
Fox	*Size*	478.0	6595.0	7654.0	Δ_{Size}	-2811.0	-3099.0	1059.0	**880.0**	1035.0	-2448.0	-1915.0
	Acc	71.26%	77.77%	77.31%	Δ_{Acc}	1.02%	1.04%	**0.45%**	0.53%	0.57%	0.88%	0.75%
Rabbit	*Size*	349.0	3196.0	5593.0	Δ_{Size}	117.0	**23.0**	2397.0	2287.0	2397.0	373.0	607.0
	Acc	69.72%	77.25%	76.41%	Δ_{Acc}	**0.21%**	0.91%	0.83%	1.0%	0.83%	0.41%	0.48%
Cat	*Size*	732.0	9589.0	11722.0	Δ_{Size}	-2242.0	-2438.0	2133.0	2133.0	2133.0	**257.0**	-723.0
	Acc	64.44%	69.68%	69.41%	Δ_{Acc}	0.47%	0.34%	**0.27%**	**0.27%**	**0.27%**	0.34%	0.4%
Panda	*Size*	351.0	3137.0	5628.0	Δ_{Size}	**-756.0**	-686.0	1199.0	1531.0	2114.0	-1246.0	-476.0
	Acc	76.09%	83.43%	82.67%	Δ_{Acc}	**0.31%**	0.4%	0.47%	0.64%	0.66%	0.88%	0.4%
Elephant	*Size*	392.0	2525.0	6282.0	Δ_{Size}	119.0	**40.0**	3757.0	2173.0	3002.0	119.0	-79.0
	Acc	79.07%	85.34%	84.8%	Δ_{Acc}	**0.35%**	0.38%	0.54%	0.64%	0.64%	0.41%	0.57%

A.2 Experiments with Expected Model Change Strategy

In this section, we present the experimental results coupled with EMC sampling strategy. The following two subsections detail the results for kernel LR and kernel SVMs, respectively.

A.2.1 Experiments for Kernel LR. Table 9-Table 10 present the results of kernel LR on UCI data sets with RBF kernel and PN kernel, respectively. Table 11-Table 12 are the results on ImageNet data sets. As shown in these tables, MS still performs the best on most cases which is consistent with previous experiments.

Table 9. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-LR with RBF kernel taking EMC as the strategy. The thresholds of MS are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Biodeg	*Size*	52.0	466.0	843.0	Δ_{Size}	**-64.0**	377.0	336.0	377.0	377.0	276.0
	Acc	82.92%	90.94%	88.77%	Δ_{Acc}	**1.7%**	2.17%	2.26%	2.17%	2.17%	1.98%
Clave	*Size*	430.0	1565.0	6880.0	Δ_{Size}	447.0	894.0	5315.0	5315.0	5315.0	**155.0**
	Acc	87.03%	98.4%	96.88%	Δ_{Acc}	**0.15%**	0.22%	1.51%	1.51%	1.51%	2.47%
Ionosphere	*Size*	17.0	123.0	280.0	Δ_{Size}	**-29.0**	-6.0	157.0	157.0	157.0	157.0
	Acc	82.25%	93.52%	87.61%	Δ_{Acc}	**3.1%**	**3.1%**	5.92%	5.92%	5.92%	5.92%
Spambase	*Size*	230.0	1324.0	3680.0	Δ_{Size}	**261.0**	2075.0	2259.0	2356.0	2356.0	2356.0
	Acc	87.92%	94.14%	92.81%	Δ_{Acc}	**0.37%**	1.31%	4.67%	1.33%	1.33%	1.33%
WDBC	*Size*	28.0	57.0	454.0	Δ_{Size}	**20.0**	32.0	323.0	384.0	394.0	-29.0
	Acc	94.26%	98.78%	97.04%	Δ_{Acc}	**1.04%**	1.57%	2.43%	1.57%	1.74%	4.52%
D-vs-P	*Size*	80.0	176.0	1286.0	Δ_{Size}	**0.0**	-22.0	996.0	1108.0	1109.0	-3.0
	Acc	83.11%	99.57%	97.02%	Δ_{Acc}	**0.12%**	0.25%	0.43%	0.5%	0.5%	0.25%
E-vs-F	*Size*	77.0	299.0	1234.0	Δ_{Size}	**-123.0**	-120.0	860.0	925.0	934.0	-222.0
	Acc	95.34%	99.61%	99.16%	Δ_{Acc}	**0.26%**	**0.26%**	0.45%	4.27%	4.27%	4.27%
M-vs-N	*Size*	78.0	163.0	1259.0	Δ_{Size}	37.0	**19.0**	965.0	1091.0	1096.0	-85.0
	Acc	95.46%	99.31%	97.26%	Δ_{Acc}	**0.05%**	**0.05%**	2.0%	1.0%	2.06%	3.85%
U-vs-V	*Size*	78.0	129.0	1260.0	Δ_{Size}	**22.0**	26.0	787.0	1131.0	1131.0	-51.0
	Acc	97.92%	99.94%	99.68%	Δ_{Acc}	**0.0%**	**0.0%**	0.25%	0.25%	0.25%	2.02%

Table 10. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-LR with PN kernel taking EMC as the strategy. The thresholds of MS are 1.0, 1.5, 1.0, 1.5, 1.0, 2.0, 2.0, 2.0, 2.0 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Biodeg	*Size*	52.0	195.0	843.0	Δ_{Size}	**275.0**	468.0	357.0	648.0	648.0	648.0
	Acc	78.77%	86.56%	84.91%	Δ_{Acc}	**0.47%**	1.18%	1.18%	1.65%	1.65%	1.65%
Clave	*Size*	430.0	1388.0	6880.0	Δ_{Size}	848.0	**160.0**	5492.0	4263.0	4902.0	885.0
	Acc	95.59%	97.98%	97.34%	Δ_{Acc}	**0.22%**	0.29%	0.64%	0.66%	0.7%	0.75%
Ionosphere	*Size*	17.0	103.0	280.0	Δ_{Size}	161.0	**116.0**	166.0	177.0	177.0	146.0
	Acc	69.53%	86.81%	83.87%	Δ_{Acc}	**2.3%**	2.56%	4.1%	4.61%	2.94%	6.02%
Spambase	*Size*	230.0	543.0	3680.0	Δ_{Size}	**-244.0**	856.0	2111.0	3137.0	3137.0	3137.0
	Acc	78.33%	85.73%	81.89%	Δ_{Acc}	**2.58%**	5.07%	5.66%	3.84%	3.84%	3.84%
WDBC	*Size*	28.0	272.0	454.0	Δ_{Size}	**-40.0**	-74.0	182.0	182.0	182.0	140.0
	Acc	85.22%	92.46%	91.3%	Δ_{Acc}	**0.29%**	2.61%	1.16%	1.16%	1.16%	1.16%
D-vs-P	*Size*	80.0	244.0	1286.0	Δ_{Size}	**-9.0**	101.0	623.0	1042.0	1042.0	997.0
	Acc	95.48%	98.59%	96.45%	Δ_{Acc}	**0.62%**	0.69%	1.38%	2.14%	2.14%	2.31%
E-vs-F	*Size*	77.0	215.0	1234.0	Δ_{Size}	**75.0**	306.0	1019.0	783.0	945.0	135.0
	Acc	95.02%	97.93%	95.79%	Δ_{Acc}	**0.58%**	0.91%	2.14%	1.75%	1.88%	1.04%
M-vs-N	*Size*	78.0	242.0	1259.0	Δ_{Size}	**30.0**	148.0	872.0	1017.0	1017.0	1017.0
	Acc	92.44%	97.27%	95.57%	Δ_{Acc}	**0.59%**	0.79%	1.62%	1.7%	1.7%	1.7%
U-vs-V	*Size*	78.0	144.0	1260.0	Δ_{Size}	**54.0**	66.0	1030.0	1116.0	1116.0	1083.0
	Acc	96.25%	98.83%	96.5%	Δ_{Acc}	**0.5%**	0.63%	2.24%	2.33%	2.33%	2.33%

Table 11. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-LR with RBF kernel taking EMC as the strategy. The thresholds of MS are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Basenji	*Size*	184.0	2774.0	2951.0	Δ_{Size}	148.0	-1443.0	-370.0	**111.0**	177.0	-1036.0
	Acc	53.45%	64.55%	61.43%	Δ_{Acc}	**0.81%**	2.71%	2.84%	3.79%	3.11%	1.62%
Chihuahua	*Size*	175.0	2275.0	2800.0	Δ_{Size}	**455.0**	-1715.0	-1050.0	490.0	525.0	-980.0
	Acc	53.29%	65.43%	65.0%	Δ_{Acc}	**0.29%**	1.0%	0.71%	1.0%	0.43%	0.71%
English setter	*Size*	242.0	3868.0	3881.0	Δ_{Size}	**-2573.0**	-1323.0	13.0	-588.0	-98.0	-637.0
	Acc	54.69%	66.01%	64.68%	Δ_{Acc}	**1.3%**	4.63%	1.34%	2.68%	3.4%	2.68%
German police dog	*Size*	174.0	2379.0	2785.0	Δ_{Size}	**35.0**	-700.0	-350.0	385.0	406.0	175.0
	Acc	58.25%	65.57%	63.27%	Δ_{Acc}	**4.02%**	4.73%	4.45%	5.6%	5.6%	7.75%
Standard poodle	*Size*	195.0	2379.0	3129.0	Δ_{Size}	**585.0**	-1014.0	-624.0	663.0	750.0	-702.0
	Acc	48.66%	76.63%	73.44%	Δ_{Acc}	**1.15%**	2.43%	2.04%	**1.15%**	3.19%	1.53%
Vizsla	*Size*	233.0	3288.0	3734.0	Δ_{Size}	**-517.0**	-2021.0	**-517.0**	282.0	282.0	-846.0
	Acc	60.92%	69.38%	67.24%	Δ_{Acc}	**2.03%**	3.0%	**2.03%**	2.36%	2.36%	2.36%
Yorkshire terrier	*Size*	304.0	4086.0	4874.0	Δ_{Size}	**-305.0**	-2379.0	-732.0	-183.0	-122.0	61.0
	Acc	49.67%	74.26%	72.87%	Δ_{Acc}	**0.33%**	1.31%	2.38%	1.97%	1.48%	1.97%
Wild dog	*Size*	242.0	2546.0	3875.0	Δ_{Size}	1296.0	-912.0	1329.0	1056.0	1152.0	**-624.0**
	Acc	70.49%	74.82%	73.99%	Δ_{Acc}	**0.31%**	0.93%	0.83%	0.62%	0.93%	0.52%
Wolf	*Size*	561.0	4257.0	8986.0	Δ_{Size}	224.0	-2623.0	**112.0**	1008.0	1568.0	560.0
	Acc	71.66%	76.33%	75.71%	Δ_{Acc}	0.27%	2.44%	0.44%	**0.22%**	0.4%	2.62%
Fox	*Size*	478.0	7006.0	7654.0	Δ_{Size}	288.0	-5664.0	-3840.0	**-1440.0**	-768.0	-2112.0
	Acc	71.94%	76.38%	72.2%	Δ_{Acc}	0.63%	1.52%	2.4%	**0.31%**	1.04%	1.1%
Rabbit	*Size*	349.0	4899.0	5593.0	Δ_{Size}	560.0	-2660.0	-1330.0	**140.0**	694.0	-280.0
	Acc	69.12%	75.98%	73.12%	Δ_{Acc}	**1.22%**	2.5%	2.93%	3.43%	2.86%	1.5%
Cat	*Size*	732.0	9405.0	11722.0	Δ_{Size}	**-5145.0**	-6468.0	-4116.0	1323.0	1617.0	-1470.0
	Acc	61.97%	68.86%	64.5%	Δ_{Acc}	**0.78%**	1.33%	0.99%	1.19%	2.52%	0.99%
Panda	*Size*	351.0	3641.0	5628.0	Δ_{Size}	1260.0	-980.0	**-630.0**	1610.0	1680.0	1987.0
	Acc	76.07%	82.32%	78.98%	Δ_{Acc}	**0.21%**	1.99%	1.78%	2.84%	2.27%	3.34%
Elephant	*Size*	392.0	3078.0	6282.0	Δ_{Size}	**1699.0**	3204.0	3204.0	3204.0	3204.0	3204.0
	Acc	80.66%	85.69%	83.81%	Δ_{Acc}	**0.83%**	1.88%	1.88%	1.88%	1.88%	1.88%

Table 12. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-LR with PN kernel taking EMC as the strategy. The thresholds of MS are all 2.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Basenji	*Size*	184.0	2487.0	2951.0	Δ_{Size}	**-496.0**	464.0	-677.0	464.0	464.0	464.0
	Acc	54.79%	60.98%	58.49%	Δ_{Acc}	**1.86%**	2.49%	3.52%	2.49%	2.49%	2.49%
Chihuahua	*Size*	175.0	2435.0	2800.0	Δ_{Size}	**-495.0**	365.0	-1025.0	365.0	365.0	365.0
	Acc	51.45%	64.33%	62.84%	Δ_{Acc}	**1.02%**	1.49%	4.41%	1.49%	1.49%	1.49%
English setter	*Size*	242.0	3390.0	3881.0	Δ_{Size}	**-70.0**	-297.0	-1225.0	491.0	491.0	491.0
	Acc	54.09%	61.15%	58.14%	Δ_{Acc}	**2.01%**	5.74%	9.68%	3.01%	3.01%	3.01%
German police dog	*Size*	174.0	2516.0	2785.0	Δ_{Size}	**184.0**	270.0	-1458.0	270.0	270.0	270.0
	Acc	56.13%	63.06%	62.34%	Δ_{Acc}	**0.47%**	0.72%	7.46%	0.72%	0.72%	0.72%
Standard poodle	*Size*	195.0	2896.0	3129.0	Δ_{Size}	-1782.0	**233.0**	-1061.0	**233.0**	**233.0**	**233.0**
	Acc	58.01%	69.19%	67.98%	Δ_{Acc}	3.29%	**1.21%**	1.72%	**1.21%**	**1.21%**	**1.21%**
Vizsla	*Size*	233.0	2959.0	3734.0	Δ_{Size}	**-423.0**	775.0	-1285.0	775.0	775.0	775.0
	Acc	53.71%	62.38%	60.85%	Δ_{Acc}	**1.46%**	1.53%	7.42%	1.53%	1.53%	1.53%
Yorkshire terrier	*Size*	304.0	4147.0	4874.0	Δ_{Size}	**647.0**	-1079.0	-1466.0	727.0	727.0	727.0
	Acc	55.05%	71.2%	70.55%	Δ_{Acc}	**0.6%**	13.09%	9.13%	0.66%	0.66%	0.66%
Wild dog	*Size*	242.0	2551.0	3875.0	Δ_{Size}	1325.0	1325.0	**-404.0**	1325.0	1325.0	1325.0
	Acc	62.07%	69.79%	69.17%	Δ_{Acc}	**0.62%**	**0.62%**	2.37%	**0.62%**	**0.62%**	**0.62%**
Wolf	*Size*	561.0	8743.0	8986.0	Δ_{Size}	-5690.0	**243.0**	**243.0**	**243.0**	**243.0**	**243.0**
	Acc	62.53%	70.2%	69.53%	Δ_{Acc}	2.32%	**0.67%**	**0.67%**	**0.67%**	**0.67%**	**0.67%**
Fox	*Size*	478.0	4606.0	7654.0	Δ_{Size}	**1680.0**	3048.0	3048.0	3048.0	3048.0	3048.0
	Acc	63.22%	70.9%	69.17%	Δ_{Acc}	**1.02%**	1.72%	1.72%	1.72%	1.72%	1.72%
Rabbit	*Size*	349.0	3359.0	5593.0	Δ_{Size}	**-490.0**	2234.0	2234.0	2234.0	2234.0	2234.0
	Acc	56.15%	68.87%	66.01%	Δ_{Acc}	**2.68%**	2.86%	2.86%	2.86%	2.86%	2.86%
Cat	*Size*	732.0	1908.0	11722.0	Δ_{Size}	**-294.0**	9814.0	9814.0	9814.0	9814.0	9814.0
	Acc	64.36%	65.11%	63.4%	Δ_{Acc}	**1.26%**	1.71%	1.71%	1.71%	1.71%	1.71%
Panda	*Size*	351.0	5563.0	5628.0	Δ_{Size}	**47.0**	65.0	-901.0	65.0	65.0	65.0
	Acc	73.11%	78.5%	78.39%	Δ_{Acc}	**0.09%**	0.12%	5.23%	0.12%	0.12%	0.12%
Elephant	*Size*	392.0	5646.0	6282.0	Δ_{Size}	**338.0**	637.0	-1817.0	637.0	637.0	637.0
	Acc	73.12%	78.91%	77.23%	Δ_{Acc}	**1.59%**	1.69%	16.03%	1.69%	1.69%	1.69%

A.2.2 Experiments for Kernel SVM. Table 13-Table 14 present the results of kernel SVM on UCI data sets with RBF kernel and PN kernel, respectively. Table 15-Table 16 show the results of kernel SVM on ImageNet data sets. As before, MS is still observed to outperform the baselines and it achieves the highest accuracy with fewer training examples on most cases.

Table 13. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-SVM with RBF kernel taking EMC as the strategy. The thresholds of MS are 2.0, 1.0, 1.5, 2.5, 2.0, 2.0, 2.0, 2.0, 1.0 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Biodeg	*Size*	52.0	320.0	843.0	Δ_{Size}	**7.0**	97.0	124.0	523.0	523.0	142.0	151.0
	Acc	77.83%	88.38%	87.19%	Δ_{Acc}	**0.99%**	1.56%	2.04%	1.19%	1.19%	1.51%	2.93%
Clave	*Size*	430.0	5031.0	6880.0	Δ_{Size}	**-2795.0**	-3712.0	-3698.0	1849.0	1849.0	-4601.0	-1063.0
	Acc	96.2%	98.31%	98.17%	Δ_{Acc}	**0.13%**	0.39%	0.21%	0.15%	0.15%	2.11%	0.25%
Ionosphere	*Size*	17.0	75.0	280.0	Δ_{Size}	50.0	55.0	76.0	200.0	205.0	198.0	**34.0**
	Acc	76.46%	94.97%	93.56%	Δ_{Acc}	**1.01%**	1.21%	2.21%	1.41%	1.41%	1.21%	10.26%
Spambase	*Size*	230.0	1741.0	3680.0	Δ_{Size}	**-434.0**	-710.0	486.0	1939.0	1939.0	1209.0	-355.0
	Acc	87.36%	93.62%	93.36%	Δ_{Acc}	**0.17%**	0.43%	0.51%	0.26%	0.26%	0.26%	1.49%
WDBC	*Size*	28.0	149.0	454.0	Δ_{Size}	302.0	-40.0	71.0	**30.0**	305.0	51.0	-70.0
	Acc	93.39%	98.0%	97.3%	Δ_{Acc}	**0.7%**	1.04%	0.83%	0.91%	**0.7%**	0.96%	2.35%
D-vs-P	*Size*	80.0	185.0	1286.0	Δ_{Size}	**-11.0**	-18.0	247.0	9.0	1086.0	-97.0	175.0
	Acc	97.12%	99.38%	99.21%	Δ_{Acc}	**0.17%**	0.19%	0.19%	0.19%	**0.17%**	1.95%	0.23%
E-vs-F	*Size*	77.0	294.0	1234.0	Δ_{Size}	**-85.0**	-93.0	119.0	677.0	939.0	940.0	23.0
	Acc	95.1%	99.55%	99.14%	Δ_{Acc}	**0.3%**	0.39%	0.32%	0.43%	0.37%	0.41%	0.8%
M-vs-N	*Size*	78.0	906.0	1259.0	Δ_{Size}	**-93.0**	-709.0	-562.0	353.0	353.0	353.0	-607.0
	Acc	94.43%	98.99%	98.54%	Δ_{Acc}	**0.38%**	1.84%	1.2%	0.44%	0.44%	0.44%	1.58%
U-vs-V	*Size*	78.0	259.0	1260.0	Δ_{Size}	420.0	**-82.0**	149.0	548.0	1001.0	-181.0	580.0
	Acc	97.41%	99.65%	99.54%	Δ_{Acc}	**0.07%**	1.54%	0.7%	0.11%	0.11%	2.24%	0.11%

Table 14. The results of Δ_{Size} and Δ_{Acc} on UCI data sets for kernel-SVM with PN kernel taking EMC as the strategy. The thresholds of MS are 1.5, 1.5, 0.5, 1.5, 2.0, 1.5, 1.5, 2.0, 1.5 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Biodeg	*Size*	52.0	433.0	843.0	Δ_{Size}	**-103.0**	-42.0	-166.0	410.0	410.0	410.0	-181.0
	Acc	78.9%	89.34%	87.64%	Δ_{Acc}	**1.07%**	1.26%	2.83%	1.7%	1.7%	1.7%	3.27%
Clave	*Size*	430.0	1763.0	6880.0	Δ_{Size}	**-275.0**	-512.0	-353.0	4390.0	5117.0	-1333.0	3861.0
	Acc	94.9%	97.67%	96.34%	Δ_{Acc}	**0.24%**	0.31%	0.29%	1.21%	1.33%	2.77%	1.1%
Ionosphere	*Size*	17.0	142.0	280.0	Δ_{Size}	**-17.0**	-10.0	-2.0	138.0	138.0	116.0	-89.0
	Acc	74.65%	91.55%	87.14%	Δ_{Acc}	**3.57%**	4.98%	5.35%	4.41%	4.41%	4.79%	9.2%
Spambase	*Size*	230.0	2290.0	3680.0	Δ_{Size}	-1126.0	-1314.0	**106.0**	1089.0	1390.0	775.0	-1603.0
	Acc	86.77%	93.0%	92.84%	Δ_{Acc}	1.19%	1.8%	1.06%	**0.15%**	0.16%	0.25%	3.18%
WDBC	*Size*	28.0	109.0	454.0	Δ_{Size}	**-49.0**	-22.0	83.0	285.0	345.0	307.0	-47.0
	Acc	92.61%	98.35%	97.13%	Δ_{Acc}	**0.87%**	1.3%	1.39%	1.22%	1.22%	1.3%	1.65%
D-vs-P	*Size*	80.0	178.0	1286.0	Δ_{Size}	**-9.0**	-11.0	286.0	30.0	1099.0	28.0	377.0
	Acc	96.19%	99.36%	99.13%	Δ_{Acc}	**0.12%**	0.21%	0.25%	0.19%	0.23%	0.19%	0.19%
E-vs-F	*Size*	77.0	181.0	1234.0	Δ_{Size}	**-2.0**	**2.0**	176.0	320.0	1050.0	1054.0	215.0
	Acc	96.21%	98.96%	98.77%	Δ_{Acc}	0.45%	0.26%	0.16%	0.16%	0.19%	0.19%	**0.13%**
M-vs-N	*Size*	78.0	246.0	1259.0	Δ_{Size}	**-16.0**	-47.0	234.0	529.0	996.0	1013.0	172.0
	Acc	95.39%	99.59%	99.25%	Δ_{Acc}	**0.25%**	0.41%	0.36%	0.29%	0.36%	0.34%	0.38%
U-vs-V	*Size*	78.0	184.0	1260.0	Δ_{Size}	**-14.0**	-35.0	179.0	513.0	1012.0	-106.0	536.0
	Acc	97.6%	99.87%	99.68%	Δ_{Acc}	**0.09%**	0.32%	0.19%	0.19%	0.16%	2.27%	0.19%

Table 15. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-SVM with RBF kernel taking EMC as the strategy. The thresholds of MS are 1.0, 1.0, 1.0, 1.5, 1.5, 1.0, 1.0, 1.0, 0.5, 0.5, 1.0, 1.0, 0.5, 1.0 respectively.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Basenji	*Size*	184.0	2355.0	2951.0	Δ_{Size}	**596.0**	**596.0**	-999.0	**596.0**	**596.0**	**596.0**	-987.0
	Acc	51.83%	66.67%	65.76%	Δ_{Acc}	**0.9%**	0.9%	3.61%	**0.9%**	**0.9%**	**0.9%**	3.25%
Chihuahua	*Size*	175.0	2188.0	2800.0	Δ_{Size}	**284.0**	346.0	-324.0	613.0	613.0	613.0	-1330.0
	Acc	58.36%	65.2%	64.3%	Δ_{Acc}	**0.88%**	0.93%	1.73%	0.89%	0.89%	0.89%	2.98%
English setter	*Size*	242.0	2982.0	3881.0	Δ_{Size}	**702.0**	857.0	-984.0	899.0	899.0	899.0	-1594.0
	Acc	53.53%	65.18%	64.48%	Δ_{Acc}	**0.59%**	0.71%	2.82%	0.7%	0.7%	0.7%	3.33%
German police dog	*Size*	174.0	1812.0	2785.0	Δ_{Size}	655.0	973.0	**-122.0**	973.0	973.0	973.0	-1057.0
	Acc	56.47%	63.87%	62.96%	Δ_{Acc}	**0.83%**	0.92%	3.85%	0.92%	0.92%	0.92%	3.79%
Standard poodle	*Size*	195.0	2353.0	3129.0	Δ_{Size}	**149.0**	217.0	-796.0	776.0	776.0	260.0	-1564.0
	Acc	65.9%	74.64%	74.02%	Δ_{Acc}	**0.38%**	0.68%	2.31%	0.62%	0.62%	0.74%	4.37%
Vizsla	*Size*	233.0	3153.0	3734.0	Δ_{Size}	581.0	**171.0**	-611.0	581.0	581.0	581.0	-2027.0
	Acc	59.64%	68.52%	68.31%	Δ_{Acc}	**0.21%**	0.43%	1.63%	**0.21%**	**0.21%**	**0.21%**	3.1%
Yorkshire terrier	*Size*	304.0	4140.0	4874.0	Δ_{Size}	**-349.0**	-609.0	-398.0	734.0	734.0	-622.0	-2348.0
	Acc	68.51%	74.84%	74.51%	Δ_{Acc}	0.65%	0.66%	0.56%	**0.33%**	**0.33%**	0.76%	2.11%
Wild dog	*Size*	242.0	2919.0	3875.0	Δ_{Size}	760.0	**-304.0**	-713.0	956.0	956.0	956.0	-1589.0
	Acc	66.86%	73.93%	73.44%	Δ_{Acc}	**0.45%**	0.86%	1.22%	0.48%	0.48%	0.48%	2.08%
Wolf	*Size*	561.0	8177.0	8986.0	Δ_{Size}	**809.0**	-2688.0	**809.0**	**809.0**	809.0	-2725.0	-4256.0
	Acc	71.71%	77.21%	77.09%	Δ_{Acc}	**0.12%**	0.55%	**0.12%**	**0.12%**	**0.12%**	0.53%	1.5%
Fox	*Size*	478.0	7195.0	7654.0	Δ_{Size}	**419.0**	-2408.0	459.0	459.0	459.0	-2227.0	-1885.0
	Acc	71.26%	77.75%	77.59%	Δ_{Acc}	**0.15%**	0.73%	0.16%	0.16%	0.16%	0.67%	0.72%
Rabbit	*Size*	349.0	4598.0	5593.0	Δ_{Size}	-693.0	-574.0	**226.0**	995.0	995.0	-595.0	-1897.0
	Acc	68.43%	76.48%	76.21%	Δ_{Acc}	0.72%	0.64%	0.62%	**0.26%**	**0.26%**	0.63%	1.73%
Cat	*Size*	732.0	10022.0	11722.0	Δ_{Size}	1700.0	-1852.0	**-1682.0**	1700.0	1700.0	1700.0	-1154.0
	Acc	64.68%	69.28%	68.92%	Δ_{Acc}	**0.36%**	0.49%	1.12%	**0.36%**	**0.36%**	**0.36%**	0.67%
Panda	*Size*	351.0	4341.0	5628.0	Δ_{Size}	**-1400.0**	-1372.0	1287.0	1287.0	1287.0	764.0	-1451.0
	Acc	77.64%	83.17%	82.77%	Δ_{Acc}	**0.34%**	0.37%	0.4%	0.4%	0.4%	0.38%	0.8%
Elephant	*Size*	392.0	4387.0	6282.0	Δ_{Size}	**-818.0**	-1151.0	1510.0	1895.0	1895.0	-450.0	-2054.0
	Acc	79.94%	85.23%	84.86%	Δ_{Acc}	**0.37%**	0.48%	0.42%	**0.37%**	**0.37%**	0.52%	0.85%

Table 16. The results of Δ_{Size} and Δ_{Acc} on ImageNet data sets for kernel-SVM with PN kernel taking EMC as the strategy. The thresholds of MS are are all 1.0.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME	QUIRE
Basenji	*Size*	184.0	2490.0	2951.0	Δ_{Size}	**407.0**	461.0	-2060.0	461.0	461.0	461.0	-691.0
	Acc	54.8%	65.36%	64.37%	Δ_{Acc}	**0.81%**	0.99%	4.51%	0.99%	0.99%	0.99%	0.9%
Chihuahua	*Size*	175.0	2076.0	2800.0	Δ_{Size}	**518.0**	725.0	-1173.0	725.0	725.0	725.0	-1022.0
	Acc	56.09%	66.41%	64.73%	Δ_{Acc}	**1.41%**	1.69%	4.63%	1.69%	1.69%	1.69%	3.4%
English setter	*Size*	242.0	3011.0	3881.0	Δ_{Size}	**821.0**	871.0	-1493.0	871.0	871.0	871.0	-1795.0
	Acc	53.22%	64.8%	63.03%	Δ_{Acc}	**1.54%**	1.78%	3.51%	1.78%	1.78%	1.78%	3.58%
German police dog	*Size*	174.0	2527.0	2785.0	Δ_{Size}	**153.0**	258.0	-1320.0	258.0	258.0	258.0	-1310.0
	Acc	55.84%	65.08%	64.16%	Δ_{Acc}	**0.53%**	0.93%	3.68%	0.93%	0.93%	0.93%	3.16%
Standard poodle	*Size*	195.0	2535.0	3129.0	Δ_{Size}	94.0	**-82.0**	-1435.0	594.0	594.0	594.0	-1377.0
	Acc	61.44%	75.71%	74.75%	Δ_{Acc}	**0.64%**	0.79%	5.24%	0.96%	0.96%	0.96%	3.68%
Vizsla	*Size*	233.0	3250.0	3734.0	Δ_{Size}	**362.0**	484.0	-1612.0	484.0	484.0	484.0	-1166.0
	Acc	59.59%	70.84%	70.02%	Δ_{Acc}	**0.71%**	0.81%	3.9%	0.81%	0.81%	0.81%	3.1%
Yorkshire terrier	*Size*	304.0	3647.0	4874.0	Δ_{Size}	775.0	-360.0	**-95.0**	1227.0	1227.0	1216.0	-1354.0
	Acc	68.63%	76.11%	75.5%	Δ_{Acc}	**0.56%**	0.87%	1.49%	0.61%	0.61%	0.66%	1.48%
Wild dog	*Size*	242.0	3015.0	3875.0	Δ_{Size}	69.0	**50.0**	-1389.0	860.0	860.0	860.0	-1074.0
	Acc	63.92%	74.96%	74.39%	Δ_{Acc}	1.82%	1.11%	3.38%	**0.58%**	**0.58%**	**0.58%**	1.98%
Wolf	*Size*	561.0	4369.0	8986.0	Δ_{Size}	4480.0	1120.0	4617.0	4617.0	4617.0	4617.0	**224.0**
	Acc	71.89%	76.89%	76.53%	Δ_{Acc}	**0.33%**	0.42%	0.36%	0.36%	0.36%	0.36%	0.58%
Fox	*Size*	478.0	7294.0	7654.0	Δ_{Size}	**-480.0**	-2400.0	360.0	360.0	360.0	360.0	-2304.0
	Acc	71.68%	77.59%	77.27%	Δ_{Acc}	**0.29%**	0.78%	0.31%	0.31%	0.31%	0.31%	0.76%
Rabbit	*Size*	349.0	4787.0	5593.0	Δ_{Size}	**420.0**	-588.0	-802.0	806.0	806.0	806.0	-1344.0
	Acc	69.09%	76.7%	76.23%	Δ_{Acc}	**0.46%**	0.86%	1.52%	0.47%	0.47%	0.47%	0.84%
Cat	*Size*	732.0	8964.0	11722.0	Δ_{Size}	2702.0	882.0	-1635.0	2758.0	2758.0	2758.0	**-221.0**
	Acc	64.51%	70.0%	69.65%	Δ_{Acc}	0.32%	**0.29%**	1.48%	0.36%	0.36%	0.36%	0.36%
Panda	*Size*	351.0	3431.0	5628.0	Δ_{Size}	1081.0	-809.0	**269.0**	2197.0	2197.0	1784.0	-1003.0
	Acc	76.63%	84.19%	83.66%	Δ_{Acc}	**0.5%**	0.52%	1.18%	0.53%	0.53%	0.51%	0.77%
Elephant	*Size*	392.0	4842.0	6282.0	Δ_{Size}	**527.0**	-2159.0	1440.0	1440.0	1440.0	-2107.0	-2370.0
	Acc	78.94%	85.43%	84.84%	Δ_{Acc}	**0.49%**	1.1%	0.59%	0.59%	0.59%	1.21%	1.21%

B EXPERIMENTS FOR NON-GRADIENT-BASED MODELS

In this section, we validate our proposed algorithm on one non-gradient-based model: the gaussian process regression (GPR) model (Section 7.2). We conduct experiments on four regression data sets: **Concrete**, **Housing**, **WineRed** from the UCI machine learning repository and **PM10** from the StatLib repository[1]. The performance of the regression model is measured by the root mean square error (RMSE) [1]. Similarly, we use Δ_{Size} and $\Delta_{Rmse} = Rmse_{\text{SC}} - Rmse_{\text{HPP}}$ to evaluate the performance of each stopping criterion.

Table 17 presents the comparison results. Our proposed MS is observed to perform the best and it obtains the lowest RMSE with fewer annotation costs, which demonstrates the effectiveness of our method on GPR model. The other methods perform not well which is similar to previous results.

Table 17. The results of Δ_{Size} and Δ_{Rmse} on UCI and StatLib data sets for GPR taking EMC as the active learning method. The thresholds of MS are all 0.1.

Dataset		Initial	HPP	All		MS	SP	GRAD	OU	MC	ME
Concrete	*Size*	51.0	746.0	823.0	Δ_{Size}	**15.0**	-95.0	77.0	55.0	75.0	77.0
	Rmse	13.4535	5.4988	5.6037	Δ_{Rmse}	**0.0347**	1.1227	0.1049	0.0989	0.1044	0.1049
Housing	*Size*	25.0	333.0	404.0	Δ_{Size}	**6.0**	-96.0	37.0	-8.0	69.0	71.0
	Rmse	7.8891	3.2776	3.3266	Δ_{Rmse}	**0.0192**	1.4685	0.9982	0.7145	0.0493	0.0489
PM10	*Size*	25.0	334.0	400.0	Δ_{Size}	**6.0**	-273.0	66.0	66.0	66.0	66.0
	Rmse	0.9443	0.8061	0.8078	Δ_{Rmse}	**0.0011**	0.065	0.0017	0.0017	0.0017	0.0017
WineRed	*Size*	79.0	963.0	1278.0	Δ_{Size}	204.0	-830.0	315.0	-884.0	-820.0	**36.0**
	Rmse	0.7987	0.6363	0.6386	Δ_{Rmse}	**0.0018**	0.2018	0.0023	0.1623	0.1521	0.0026

REFERENCES

[1] W. Cai, M. Zhang, and Y. Zhang. 2016. Batch mode active learning for regression with expected model change. *IEEE Transactions on Neural Networks and Learning Systems* (2016).

[1] http://lib.stat.cmu.edu/datasets/