

Table 8: RQ3: Failures sensitivity analysis (confidence/lsa/dsa)

	<i>technique</i>	<i>minimum_mean</i>	<i>delta_100/50</i>	<i>delta_200/100</i>	<i>delta_400/200</i>	<i>delta_800/400</i>	<i>delta_800/50</i>	<i>maximum_mean</i>
MNIST	GBS	5.3/5.8/4.5	2.4/1.8/2.0	2.2/2.0/2.3	2.3/2.1/2.2	2.2/2.0/2.2	26.3/15.7/21.9	136.5/91.9/95.9
	DeepEST	19.3/8.0/17.9	2.1/2.2/2.1	2.0/2.1/2.0	2.0/1.9/2.0	2.0/2.0/2.0	16.7/17.7/16.5	321.6/140.0/295.9
	2-UPS	17.4/5.8/5.9	2.0/2.1/2.1	1.9/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.9/17.1/16.9	277.2/100.5/98.0
	RHC-S	17.1/6.7/6.2	2.0/1.9/1.9	2.0/2.0/2.0	1.9/2.0/2.1	2.0/2.0/2.1	16.1/15.1/16.0	274.2/101.8/100.3
	SSRS	10.0/6.9/5.5	2.0/2.1/2.1	2.0/1.9/1.9	2.0/2.0/2.1	2.0/2.0/2.0	15.5/15.8/16.6	153.3/107.4/90.9
	SUPS	17.7/6.2/5.8	2.0/2.1/2.3	2.0/1.9/1.9	2.0/2.0/2.0	2.0/2.0/2.0	16.0/16.4/17.3	282.6/102.6/100.9
	CES	3.8	2.0	2.0	2.0	2.0	16.1	61.5
	SRS	3.8	1.9	2.0	2.0	2.0	15.9	58.6
dataset CIFAR10	GBS	15.4/14.9/14.4	2.2/1.9/2.1	2.0/2.0/2.1	2.0/2.0/2.0	2.0/2.0/2.0	17.9/15.9/18.3	272.5/238.7/261.1
	DeepEST	26.9/17.1/25.3	2.0/2.0/2.1	2.0/2.1/2.0	2.0/2.0/2.0	2.0/2.1/2.0	16.1/16.2/16.2	432.7/277.7/408.6
	2-UPS	26.7/15.8/16.8	2.0/2.0/2.0	2.0/1.9/2.0	2.0/2.0/2.0	2.0/2.0/2.0	16.1/16.1/15.8	431.8/252.7/264.2
	RHC-S	27.2/15.9/16.1	2.0/2.0/2.0	1.9/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.7/16.0/16.3	427.7/252.6/262.4
	SSRS	19.7/15.3/14.4	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	16.0/15.7/16.0	315.1/239.4/231.1
	SUPS	26.7/15.1/16.4	2.0/2.1/2.0	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	16.2/16.6/16.2	433.8/250.6/263.3
	CES	14.1	2.0	2.0	1.9	2.0	15.2	216.5
	SRS	13.8	2.0	2.0	2.0	2.0	16.4	227.6
CIFAR100	GBS	21.5/21.3/21.7	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.9/15.9/16.5	341.0/339.2/357.8
	DeepEST	33.7/29.8/34.9	2.0/2.0/1.9	2.0/2.0/2.0	2.0/2.0/1.8	2.0/2.0/1.6	16.2/16.0/11.3	546.6/475.3/393.3
	2-UPS	35.7/27.2/24.5	2.0/2.0/2.0	2.0/2.1/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.9/16.0/15.7	565.9/434.7/385.8
	RHC-S	35.3/27.6/24.1	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.9/15.7/15.9	560.9/432.6/383.3
	SSRS	28.2/22.7/20.8	2.0/2.0/2.0	2.0/2.1/2.0	2.0/2.0/2.0	2.0/2.0/2.0	15.5/16.5/15.9	436.5/374.5/329.8
	SUPS	35.2/27.4/23.8	2.0/2.0/2.0	2.0/2.0/2.1	2.0/2.0/2.0	2.0/2.0/2.0	16.2/16.2/16.4	571.2/441.1/389.7
	CES	19.7	1.9	1.9	1.9	2.0	13.8	270.5
	SRS	20.8	1.9	2.0	2.0	2.0	15.2	315.0

Table 9: RQ3: Failures sensitivity analysis (lsa/VAE/SAE)

	<i>technique</i>	<i>minimum_mean</i>	<i>delta_100/50</i>	<i>delta_200/100</i>	<i>delta_400/200</i>	<i>delta_800/400</i>	<i>delta_800/50</i>	<i>maximum_mean</i>
dataset Udacity	GBS	4.2/3.7/4.1	1.5/2.1/1.9	1.6/2.8/2.0	1.2/2.7/1.9	1.0/2.1/2.0	3.0/34.1/14.9	12.4/122.8/57.8
	DeepEST	6.8/3.6/3.6	2.1/1.8/2.2	2.0/2.4/2.1	2.0/1.9/2.0	2.0/2.0/2.0	16.2/16.6/18.0	109.3/60.3/61.2
	2-UPS	7.4/3.8/3.4	2.2/1.9/2.4	1.9/2.0/2.0	2.0/2.0/1.9	2.0/2.0/2.1	16.6/15.5/18.9	123.1/59.1/61.2
	RHC-S	8.1/3.5/3.7	1.8/2.3/2.3	2.1/1.8/1.8	1.9/2.1/2.0	1.8/2.0/2.0	13.8/17.3/16.4	112.1/60.8/61.5
	SSRS	7.5/4.2/4.2	2.0/2.0/1.8	2.0/1.9/2.2	2.0/2.0/2.0	2.0/2.0/2.0	15.6/15.1/15.2	117.5/64.2/64.9
	SUPS	8.1/3.7/3.8	2.0/2.0/2.0	2.1/2.1/2.0	2.0/2.0/1.9	2.0/2.0/2.1	15.9/16.5/16.2	127.3/60.6/61.1
	CES	3.1	2.0	2.3	2.1	2.1	20.9	65.3
	SRS	3.3	2.1	2.0	2.2	2.0	17.9	58.3