

Table 3: RQ2.1: Number of failures in the classification task

		MNIST				CIFAR10				CIFAR100			
χ	technique	mean	max	min	std	mean	max	min	std	mean	max	min	std
confidence	GBS	28.20	51.00	16.00	7.23	68.42	84.00	48.00	8.04	86.17	100.00	66.00	10.59
	DeepEST	80.50	103.00	54.00	10.33	108.70	131.00	91.00	9.38	136.44	149.00	119.00	6.56
	2-UPS	69.50	98.00	37.00	17.40	108.92	136.00	86.00	11.67	141.48	156.00	125.00	6.15
	RHC-S	70.56	99.00	40.00	16.47	106.00	135.00	84.00	12.83	142.28	158.00	127.00	6.40
	SSRS	38.44	55.00	17.00	10.24	78.70	107.00	59.00	13.52	109.24	125.00	95.00	6.19
	SUPS	69.82	96.00	38.00	16.93	106.90	136.00	83.00	12.31	143.56	154.00	132.00	5.47
	CES	15.64	32.00	5.00	5.78	55.76	86.00	29.00	12.63	70.42	89.00	55.00	7.48
	SRS	14.58	29.00	3.00	5.32	57.34	84.00	30.00	13.12	77.99	104.00	10.00	12.82
LSA	GBS	21.00	31.00	8.00	5.24	58.19	80.00	36.00	10.59	84.62	105.00	66.00	7.37
	DeepEST	35.89	57.00	17.00	10.65	69.23	97.00	49.00	10.35	119.89	147.00	98.00	12.36
	2-UPS	25.04	40.00	9.00	7.45	61.80	84.00	41.00	11.26	110.46	149.00	84.00	20.11
	RHC-S	25.51	42.00	10.00	7.13	62.94	98.00	42.00	11.41	110.30	149.00	85.00	19.33
	SSRS	27.30	44.00	14.00	6.40	60.09	82.00	40.00	10.68	93.16	117.00	72.00	10.73
	SUPS	25.39	38.00	7.00	6.80	63.46	89.00	42.00	11.24	110.17	152.00	87.00	21.65
	CES	15.64	32.00	5.00	5.78	55.76	86.00	29.00	12.63	70.42	89.00	55.00	7.48
	SRS	14.58	29.00	3.00	5.32	57.34	84.00	30.00	13.12	77.99	104.00	10.00	12.82
DSA	GBS	20.27	31.00	9.00	6.00	63.57	88.00	42.00	10.97	88.50	104.00	66.00	7.48
	DeepEST	72.98	111.00	39.00	17.91	102.37	122.00	85.00	8.87	136.70	151.00	123.00	4.93
	2-UPS	25.18	37.00	8.00	7.19	65.16	89.00	38.00	11.98	96.31	116.00	77.00	8.66
	RHC-S	23.94	37.00	9.00	7.02	65.48	100.00	37.00	13.79	96.88	122.00	80.00	9.64
	SSRS	21.69	34.00	8.00	5.57	58.30	82.00	34.00	11.96	82.38	99.00	69.00	6.40
	SUPS	24.72	39.00	11.00	6.87	65.71	87.00	37.00	12.09	97.28	127.00	76.00	10.67
	CES	15.64	32.00	5.00	5.78	55.76	86.00	29.00	12.63	70.42	89.00	55.00	7.48
	SRS	14.58	29.00	3.00	5.32	57.34	84.00	30.00	13.12	77.99	104.00	10.00	12.82