Experiments design

1 Experiments
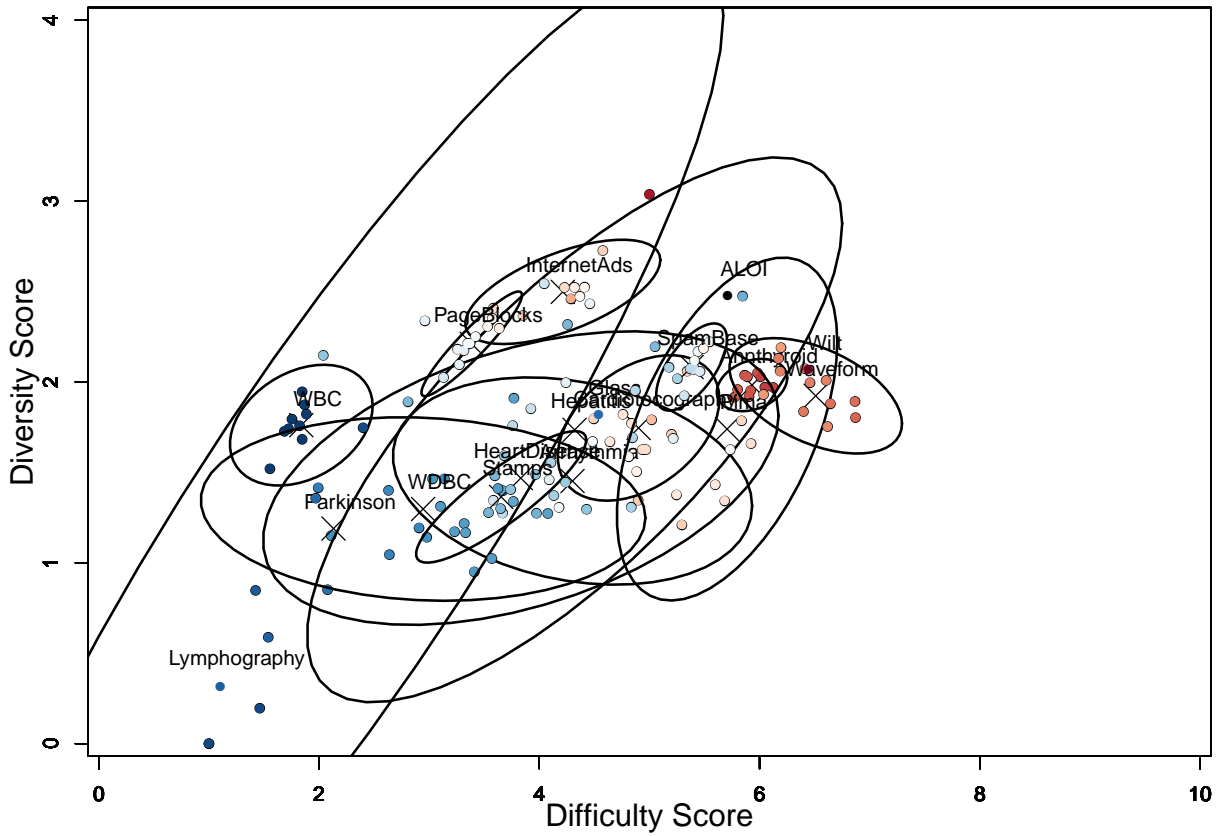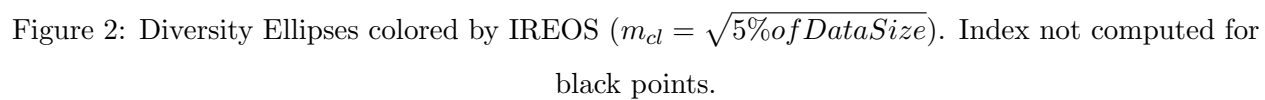
1.1 Dataset evaluation experiments

Figure 1: Diversity Ellipses colored by IREOS ($m_{cl} = 1$). Index not computed for black points.

Figure 2: Diversity Ellipses colored by IREOS ($m_{cl} = \sqrt{5\% of DataSize}$). Index not computed for black points.
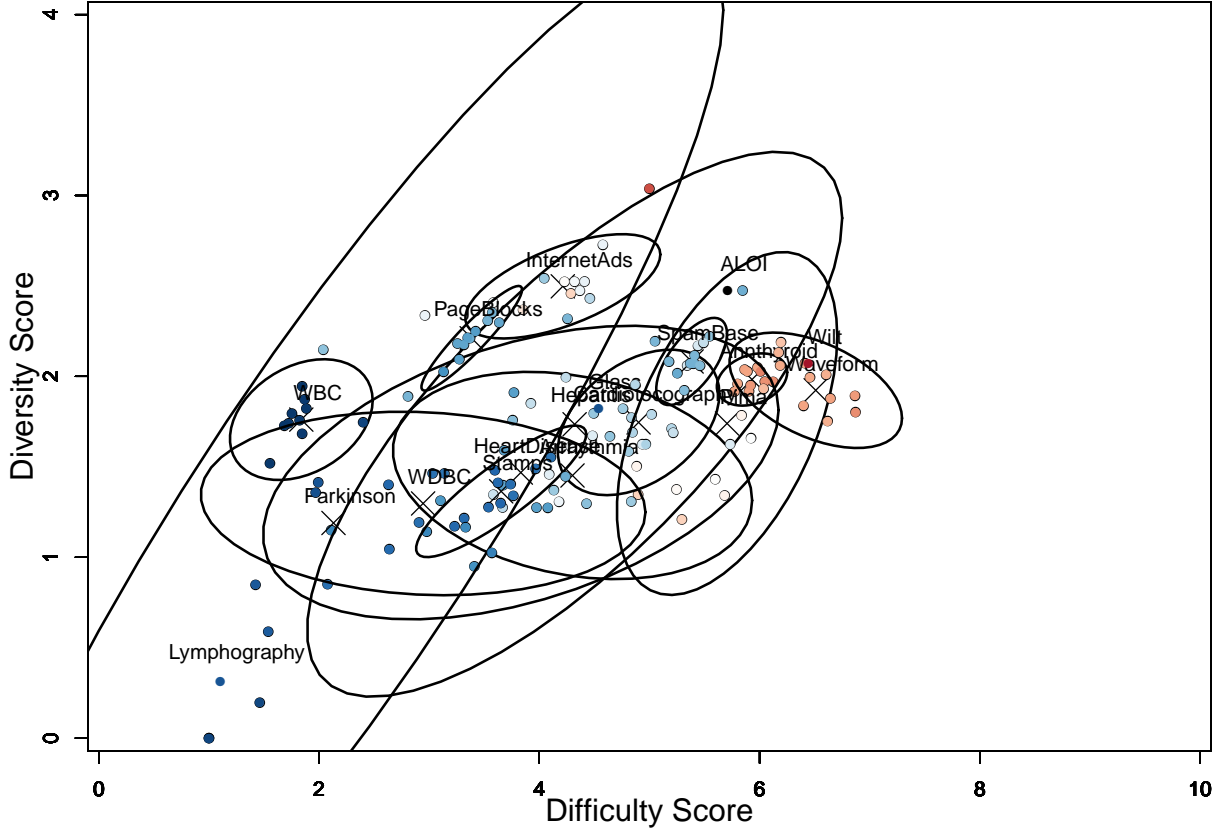
Figure 3: Diversity Ellipses colored by IREOS ($m_{cl} = n$). Index not computed for black points.

1.2 Correlation experiments

Table 1: Spearman correlation

Dataset	$m_{cl} = 1$	$m_{cl} = \sqrt{}$	$m_{cl} = n$
Annthyroid	-0.079	-0.297	-0.321
Arrhythmia	0.588	0.697	0.697
Cardiotocography	0.685	0.976	0.988
Glass	0.467	0.467	0.406
HeartDisease	0.952	0.952	0.964
Hepatitis	0.939	0.927	0.927
InternetAds	0.6	0.661	0.818
Lymphography	0.782	0.83	0.83
PageBlocks	0.976	1	0.988
Parkinson	0.6	0.6	-0.321
Pima	0.745	0.952	0.855
SpamBase	0.685	0.758	0.758
Stamps	0.842	0.915	0.879
Waveform	0.588	0.818	0.818
WBC	-0.467	-0.382	-0.418
WDBC	0.309	0.309	0.309
Wilt	-0.83	-0.927	-0.927

Table 2: Spearman correlation Synthetic datasets

Dataset	$m_{cl} = 1$	$m_{cl} = \sqrt{}$	$m_{cl} = n$
gaussian20dim_4clusters_nr1	1		
gaussian20dim_6clusters_nr1	1		
gaussian22dim_5clusters_nr1	1		
gaussian22dim_6clusters_nr1	1		
gaussian22dim_9clusters_nr1	1		
gaussian23dim_4clusters_nr1	1		
gaussian23dim_9clusters_nr1	1		
gaussian24dim_2clusters_nr1	1		
gaussian24dim_3clusters_nr1	1		
gaussian24dim_4clusters_nr1	1		
gaussian24dim_7clusters_nr1	1		
gaussian25dim_5clusters_nr1	1		
gaussian25dim_9clusters_nr1	1		
gaussian26dim_4clusters_nr1	1		
gaussian27dim_5clusters_nr1	1		
gaussian27dim_6clusters_nr1	1		
gaussian28dim_4clusters_nr1	1		
gaussian28dim_7clusters_nr1	1		
gaussian29dim_3clusters_nr1	1		
gaussian30dim_4clusters_nr1	1		
gaussian31dim_4clusters_nr1	1		
gaussian33dim_3clusters_nr1	1		
gaussian33dim_5clusters_nr1	1		
gaussian35dim_6clusters_nr1	1		
gaussian36dim_8clusters_nr1	1		
gaussian36dim_9clusters_nr1	1		
gaussian37dim_3clusters_nr1	1		
gaussian38dim_9clusters_nr1	1		
gaussian39dim_2clusters_nr1	1		
gaussian39dim_5clusters_nr1	1		

1.3 Model selection experiments

Table 3: Summarization of the ROC AUC values for all candidate solutions used in the experiments. Two last columns indicate the solutions selected by ext-IREOS for $m_{cl} = 1$ and $m_{cl} = n$ respectively

Annthyroid	0.528	0.732	0.63	0.642	KNNW	0.596	LOF	0.551	KNN
Arrhythmia	0.5	0.916	0.737	0.879	GLOSH	0.879	GLOSH	0.879	GLOSH
Cardiotocography	0.536	0.839	0.689	0.74	LOF	0.806	KNN	0.806	KNN
Glass	0.492	0.904	0.699	0.539	GLOSH	0.539	GLOSH	0.539	GLOSH
HeartDisease	0.387	0.97	0.687	0.847	KNNW	0.847	KNNW	0.909	KNNW
Hepatitis	0.075	0.96	0.522	0.871	FastABOD	0.96	COF	0.96	COF
InternetAds	0.413	0.895	0.656	0.843	FastABOD	0.843	FastABOD	0.843	FastABOD
Lymphography	0.43	1	0.748	0.847	KDEOS	1	COF	1	COF
PageBlocks	0.506	0.943	0.728	0.896	LOF	0.943	LDF	0.943	LDF
Parkinson	0.448	1	0.765	1	FastABOD	1	FastABOD	0.625	KDEOS
Pima	0.499	0.834	0.67	0.762	LOF	0.762	LOF	0.762	LOF
SpamBase	0.463	0.779	0.625	0.779	KNNW	0.779	KNNW	0.779	KNNW
Stamps	0.38	0.921	0.655	0.921	KNN	0.921	KNN	0.921	KNN
Waveform	0.5	0.796	0.66	0.708	INFLO	0.796	LDF	0.796	LDF
WBC	0.438	0.997	0.721	0.755	LDOF	0.755	LDOF	0.755	LDOF
WDBC	0.497	0.988	0.745	0.884	LOF	0.884	LOF	0.884	LOF
Wilt	0.34	0.713	0.527	0.424	KNNW	0.424	KNNW	0.424	KNNW

1.4 Computational experiments

1.4.1 Minimum number of γ

1.4.2 Separability of the minimum number of observation

1.4.3 Speed classifier up

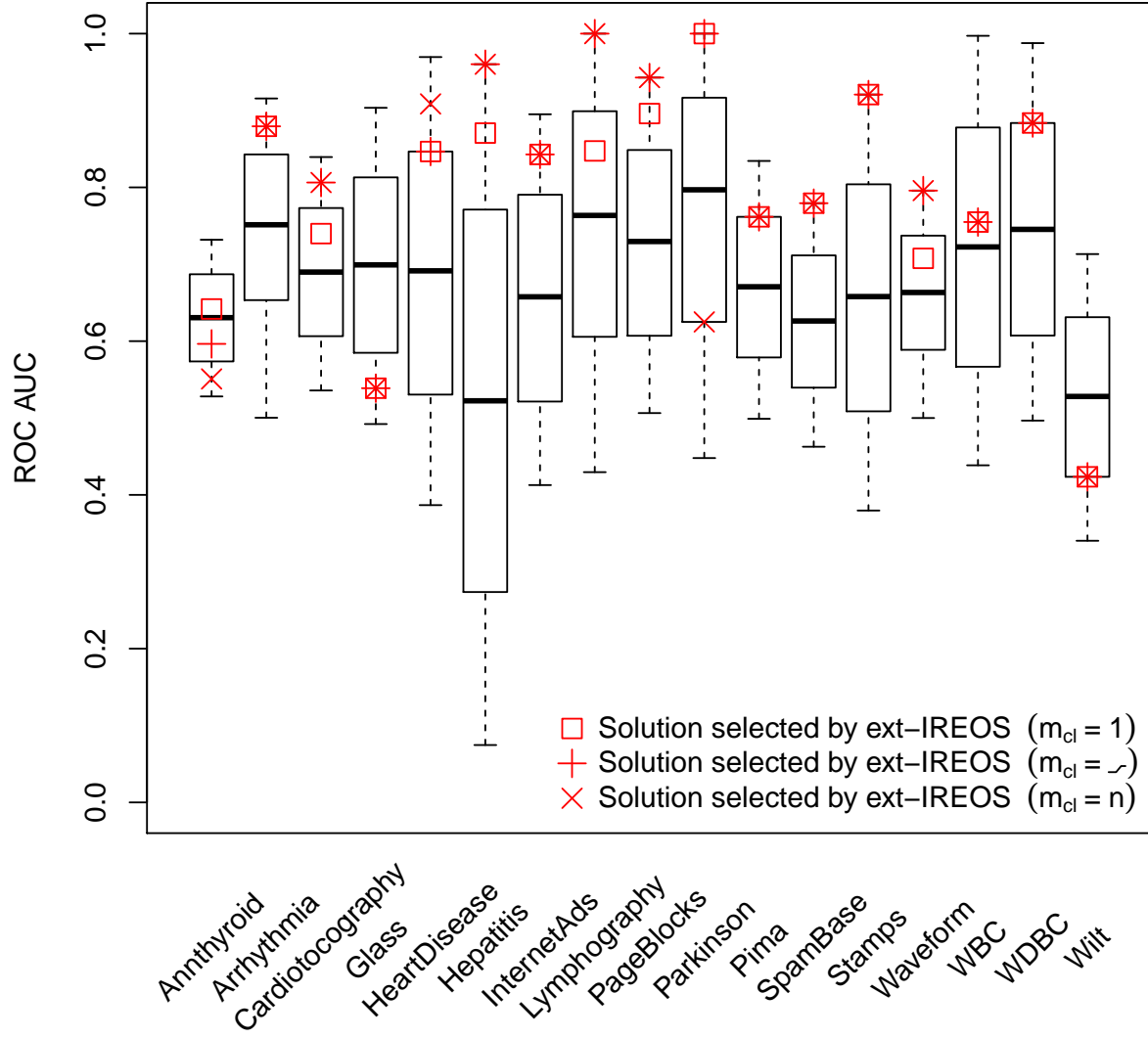Figure 4: Distribution of the ROC AUC values for all candidate solutions used in the experiments.

Table 4: Adaptive

Dataset	Original		0.01		
	Total	Runtime	Total	Runtime	Diff
Arrhythmia	25600	0:13:56	4555.6 (17.8%)	0:2:43 (19.56%)	0.00790
Cardiotocography	173400	18:26:45	13881.8 (8.01%)	1:34:39 (8.55%)	0.01950
Glass	21400	0:2:9	2978.4 (13.92%)	0:0:34 (26.39%)	0.00320
HeartDisease	15700	0:1:0	1796.2 (11.44%)	0:0:19 (32.41%)	0.00813
Hepatitis	7000	0:0:7	1130.2 (16.15%)	0:0:2 (31.81%)	0.00517
Lymphography	14800	0:0:55	1727.4 (11.67%)	0:0:16 (30.11%)	0.00683
PageBlocks	513900	491:47:43	36811.8 (7.16%)	35:9:6 (7.15%)	0.00361
Parkinson	5000	0:0:3	740.4 (14.81%)	0:0:1 (28.92%)	0.00341
Pima	52600	0:30:43	5770 (10.97%)	0:4:32 (14.76%)	0.01036
SpamBase	266100	124:9:0	35113 (13.2%)	17:59:15 (14.49%)	0.00938
Stamps	32500	0:7:3	3228.4 (9.93%)	0:0:54 (12.77%)	0.01213
Waveform	344300	157:37:13	39899 (11.59%)	21:34:18 (13.69%)	0.01692
WBC	22300	0:2:32	3362.2 (15.08%)	0:0:34 (22.29%)	0.00519
WDBC	36700	0:13:19	5017.4 (13.67%)	0:2:5 (15.68%)	0.00887
Wilt	481900	359:23:27	34881.4 (7.24%)	26:58:52 (7.51%)	0.01377

Table 5: Adaptive

Dataset	0.005			0.001		
	Total	Runtime	Diff	Total	Runtime	Diff
Arrhythmia	6104.8 (23.85%)	0:3:32 (25.38%)	0.00266	8452.6 (33.02%)	0:4:44 (34.04%)	0.00069
Cardiotocography	18755.6 (10.82%)	2:8:27 (11.61%)	0.01734	29482 (17%)	3:22:27 (18.29%)	0.01609
Glass	3765 (17.59%)	0:0:38 (29.95%)	0.00152	5239.8 (24.49%)	0:0:47 (36.65%)	0.00028
HeartDisease	2396.2 (15.26%)	0:0:21 (36.52%)	0.00352	3536 (22.52%)	0:0:26 (44.32%)	0.00072
Hepatitis	1465.8 (20.94%)	0:0:2 (36.66%)	0.00227	1987.2 (28.39%)	0:0:3 (43.19%)	0.00041
Lymphography	2330.8 (15.75%)	0:0:19 (34.55%)	0.00358	3448 (23.3%)	0:0:23 (42.92%)	0.00097
PageBlocks	40901.8 (7.96%)	39:2:50 (7.94%)	0.00199	50493.2 (9.83%)	48:9:0 (9.79%)	5e-04
Parkinson	957.6 (19.15%)	0:0:1 (32.74%)	0.00165	1374 (27.48%)	0:0:1 (39.21%)	0.00045
Pima	8111.6 (15.42%)	0:5:55 (19.29%)	0.00757	12057.4 (22.92%)	0:8:17 (26.99%)	0.00635
SpamBase	47637.8 (17.9%)	24:13:14 (19.51%)	0.00363	65931 (24.78%)	33:14:10 (26.77%)	0.00081
Stamps	4312.4 (13.27%)	0:1:7 (15.86%)	0.01015	6460.6 (19.88%)	0:1:33 (22.06%)	0.00925
Waveform	64058 (18.61%)	34:46:58 (22.07%)	0.00512	90128.4 (26.18%)	49:3:33 (31.13%)	0.00177
WBC	4414.4 (19.8%)	0:0:41 (26.85%)	0.00389	6051.2 (27.14%)	0:0:51 (33.86%)	0.00344
WDBC	7149 (19.48%)	0:2:49 (21.2%)	0.00327	10172.2 (27.72%)	0:3:50 (28.84%)	0.00207
Wilt	39469 (8.19%)	30:35:16 (8.51%)	0.01724	50346.2 (10.45%)	39:10:45 (10.9%)	0.01141

Table 6: Stopping

Dataset	Original		Weights 0		Stopping	
	Total	Runtime	Total	Runtime	Total	Runtime
Arrhythmia	25600	0:13:56	11700 (45.7%)	0:6:11 (44.42%)	11233.33 (43.88%)	0:5:56 (42.64%)
Cardiotocography	173400	18:26:45	86600 (49.94%)	9:14:12 (50.08%)	76576.67 (44.16%)	8:10:6 (44.28%)
Glass	21400	0:2:9	9570 (44.72%)	0:0:55 (42.6%)	8943.333 (41.79%)	0:0:51 (39.85%)
HeartDisease	15700	0:1:0	7800 (49.68%)	0:0:30 (51.12%)	7206.667 (45.9%)	0:0:28 (47.44%)
Hepatitis	7000	0:0:7	3400 (48.57%)	0:0:3 (48.09%)	3163.333 (45.19%)	0:0:3 (44.65%)
InternetAds	168200	1038:51:52	83970 (49.92%)	520:12:11 (50.07%)	80120 (47.63%)	496:25:4 (47.78%)
Lymphography	14800	0:0:55	6490 (43.85%)	0:0:24 (44.1%)	6066.667 (40.99%)	0:0:22 (41.41%)
PageBlocks	513900	491:47:43	234440 (45.62%)	224:9:21 (45.58%)	189413.3 (36.86%)	181:18:38 (36.87%)
Parkinson	5000	0:0:3	2180 (43.6%)	0:0:1 (41.67%)	2033.333 (40.67%)	0:0:1 (38.98%)
Pima	52600	0:30:43	23330 (44.35%)	0:13:37 (44.37%)	21966.67 (41.76%)	0:12:49 (41.78%)
SpamBase	266100	124:9:0	132400 (49.76%)	61:47:29 (49.77%)	123600 (46.45%)	57:45:50 (46.53%)
Stamps	32500	0:7:3	15900 (48.92%)	0:3:28 (49.1%)	14353.33 (44.16%)	0:3:7 (44.33%)
Waveform	344300	157:37:13	154410 (44.85%)	70:57:10 (45.01%)	149543.3 (43.43%)	68:46:1 (43.63%)
WBC	22300	0:2:32	9990 (44.8%)	0:1:5 (43.03%)	9610 (43.09%)	0:1:3 (41.47%)
WDBC	36700	0:13:19	16390 (44.66%)	0:5:54 (44.35%)	15413.33 (42%)	0:5:33 (41.72%)
Wilt	481900	359:23:27	240430 (49.89%)	179:14:18 (49.87%)	208526.7 (43.27%)	155:36:28 (43.3%)

Table 7: Nearest neighbor

Dataset	Original	10		50		100	
	Runtime	Diff	Runtime	Diff	Runtime	Diff	Runtime
Cardiotocography	18:26:45	0.05651 ± 0.07093	0:0:11	0.01953 ± 0.02443	0:1:40	0.01725 ± 0.02395	0:4:54
InternetAds	1038:51:52	0.03822 ± 0.07256	0:1:16	0.01213 ± 0.0165	0:19:40	0.01028 ± 0.00964	1:24:37
PageBlocks	491:47:43	0.09086 ± 0.10938	0:0:20	0.04187 ± 0.06783	0:3:18	0.03835 ± 0.05949	0:10:51
SpamBase	124:9:0	0.03691 ± 0.03992	0:0:15	0.02928 ± 0.02126	0:3:5	0.03367 ± 0.03464	0:11:16
Waveform	157:37:13	0.03083 ± 0.01315	0:0:11	0.01309 ± 0.00637	0:2:44	0.00786 ± 0.00465	0:9:7
Wilt	359:23:27	0.11055 ± 0.13974	0:0:17	0.05106 ± 0.05342	0:2:47	0.04109 ± 0.03445	0:9:4

Table 8: Nearest neighbor

Dataset	Original	Combined approaches	
	Runtime	Runtime	Diff
Arrhythmia	0:13:56	0:0:14 (1.79%)	0.00459
Cardiotocography	18:26:45	0:0:17 (0.03%)	0.00869
Glass	0:2:9	0:0:3 (2.96%)	0.00156
HeartDisease	0:1:0	0:0:3 (5.55%)	0.00282
Hepatitis	0:0:7	0:0:1 (17.34%)	0.00222
InternetAds	1038:51:52	0:13:19 (0.02%)	0.01042
Lymphography	0:0:55	0:0:2 (4.78%)	0.00337
PageBlocks	491:47:43	0:0:24 (0%)	0.0018
Parkinson	0:0:3	0:0:0 (16.96%)	0.00187
Pima	0:30:43	0:0:7 (0.42%)	0.00705
Stamps	0:7:3	0:0:4 (1.07%)	0.0035
SpamBase	124:9:0	0:1:34 (0.02%)	0.02389
WBC	0:2:32	0:0:4 (2.92%)	0.00464
WDBC	0:13:19	0:0:7 (0.88%)	0.00658
Wilt	359:23:27	0:0:25 (0%)	0.00316