

# On developing an automatic threshold applied to feature selection ensembles

Borja Seijo-Pardo, Verónica Bolón-Canedo, and Amparo Alonso-Betanzos

Department of Computer Science - University of A Coruña  
Campus de Elviña s/n 15071 - A Coruña, Spain

## A Appendix: Supplementary Material

This appendix shows the detailed results of the several experiments conducted in this study.

**Table 1.** Average percentage test classification error (T-error) and number of selected features (#feats) for classical datasets.

		Spambase		Madelon		Connect4		Isolet		USPS		
		T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats	
FIXED TH	DESIGN CT	100%	19.104	57	44.375	500	34.170	42	25.420	617	4.936	256
		50%	19.082	29	35.458	250	34.156	21	24.689	309	5.711	128
		25%	20.235	14	34.042	125	33.706	11	29.512	154	7.475	64
		10%	26.190	6	34.167	50	34.014	4	39.644	62	15.982	26
		log2(n)	26.190	6	32.333	9	33.188	5	67.667	9	33.824	8
AUTOMATIC THRESHOLD	DESIGN CT	F1MIN	25.125	9	32.333	9	33.398	7	59.458	14	19.219	21
		F2MIN	26.190	6	32.333	9	33.188	5	67.667	9	33.824	8
		F3MIN	26.190	6	32.333	9	33.188	5	67.667	9	33.824	8
	DESIGN TC	F1U1	18.039	36	33.792	27	33.908	16	44.684	37	7.389	52
		F1U2	17.822	31	31.792	11	31.157	8	58.253	17	20.327	13
		F1U3	17.778	30	31.792	11	31.251	7	65.936	10	27.06	9
		F1U4	17.865	30	31.792	10	31.403	6	68.218	9	31.104	8
		F1U5	17.887	30	31.958	9	31.772	5	68.218	9	31.104	8
		F1U6	17.887	30	31.958	9	31.772	5	68.218	9	31.104	8
		F2U1	19.061	19	33.792	27	33.908	16	44.684	37	10.131	39
		F2U2	19.061	12	31.792	11	31.157	8	58.253	17	25.554	11
		F2U3	23.604	10	31.792	11	31.251	7	65.936	10	31.104	8
		F2U4	25.212	9	31.792	10	31.403	6	68.218	9	31.104	8
		F2U5	29.689	6	31.958	9	31.772	5	68.218	9	31.104	8
		F2U6	29.689	6	31.958	9	31.772	5	68.218	9	31.104	8
		F3U1	19.061	19	33.792	27	33.908	16	44.684	37	10.131	39
		F3U2	19.061	12	31.792	11	31.157	8	58.253	17	25.554	11
		F3U3	23.604	10	31.792	11	31.251	7	65.936	10	31.104	8
		F3U4	25.212	9	31.792	10	31.403	6	68.218	9	31.104	8
		F3U5	29.689	6	31.958	9	31.772	5	68.218	9	31.104	8
		F3U6	29.689	6	31.958	9	31.772	5	68.218	9	31.104	8
	CFL3	19.887	14	32.042	13	31.251	7	60.626	19	33.513	22	

**Table 2.** Fitness measure  $F$  results for classical datasets.

		Spambase	Madelon	Connect4	Isolet	USPS		
FIXED TH	DESIGN CT	100%	13.0690	29.9170	23.1130	17.2800	3.6243	
		50%	12.8910	23.8050	22.9370	16.6260	3.9740	
		25%	13.5720	22.7780	22.5580	19.7580	5.0666	
		10%	17.4950	22.8110	22.7080	26.4630	10.6890	
		log2(n)	17.4950	21.5610	22.1650	45.1160	22.5600	
	AUTOMATIC THRESHOLD	DESIGN CT	F1MIN	16.8030	21.5610	22.3210	39.6460	12.8400
			F2MIN	17.4950	21.5610	22.1650	45.1160	22.5600
			F3MIN	17.4950	21.5610	22.1650	45.1160	22.5600
		DESIGN TC	F1U1	12.2370	22.5460	22.7320	29.8090	4.9936
			F1U2	12.0630	21.2020	20.8350	38.8450	13.5680
			F1U3	12.0270	21.2020	20.8900	43.9630	18.0520
			F1U4	12.0850	21.2010	20.9830	45.4840	20.7460
			F1U5	12.1000	21.3110	21.2210	45.4840	20.7460
			F1U6	12.1000	21.3110	21.2210	45.4840	20.7460
			F2U1	12.8180	22.5460	22.7320	29.8090	6.8048
			F2U2	12.7780	21.2020	20.8350	38.8450	17.0500
			F2U3	15.7940	21.2020	20.8900	43.9630	20.7460
			F2U4	16.8610	21.2010	20.9830	45.4840	20.7460
			F2U5	19.8280	21.3110	21.2210	45.4840	20.7460
			F2U6	19.8280	21.3110	21.2210	45.4840	20.7460
F3U1			12.8180	22.5460	22.7320	29.8090	6.8048	
F3U2			12.7780	21.2020	20.8350	38.8450	17.0500	
F3U3			15.7940	21.2020	20.8900	43.9630	20.7460	
F3U4			16.8610	21.2010	20.9830	45.4840	20.7460	
F3U5			19.8280	21.3110	21.2210	45.4840	20.7460	
F3U6	19.8280	21.3110	21.2210	45.4840	20.7460			
CFL3	13.3400	21.3700	20.8900	40.4280	22.3710			

**Table 3.** Average percentage test classification error (T-error) and number of selected features (#feats) for microarray datasets.

		Colon		DLBCL		Leukemia		Lung		Ovarian		11 Tumors		Leukemia 2			
		T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats	T-error	#feats		
AUTOMATIC THRESHOLD	FIXED TH	DESIGN CT	100%	40.000	2000	20.000	4026	41.176	7129	4.698	12533	38.095	15154	86.207	12533	54.167	11225
			50%	40.000	1000	13.333	2013	41.176	3565	1.342	6267	9.524	7577	84.483	6267	25.000	5613
			25%	40.000	500	13.333	1007	41.176	1782	0.671	3133	7.143	3789	68.966	3133	25.000	2806
			10%	40.000	200	13.333	403	41.176	713	0.671	1253	4.762	1515	68.966	1253	25.000	1123
			log2(n)	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13
			F1MIN	40.000	231	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13
	F2MIN	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13		
	F3MIN	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	68.966	28	12.500	13		
	DESIGN TC	F1U1	35.000	96	0.000	47	17.647	51	1.342	50	0.000	35	68.966	84	8.333	64	
		F1U2	35.000	78	6.667	29	20.588	22	0.671	22	0.000	22	70.690	19	8.333	25	
		F1U3	25.000	20	13.333	15	11.765	14	0.671	15	1.190	16	77.586	14	4.167	15	
		F1U4	15.000	12	13.333	13	14.706	13	0.671	14	1.190	16	77.586	14	4.167	13	
		F1U5	15.000	11	13.333	13	14.706	13	0.671	14	2.380	14	77.586	14	4.167	13	
		F1U6	15.000	11	13.333	12	14.706	13	0.671	14	2.381	14	77.586	14	4.167	13	
		F2U1	25.000	36	13.333	33	17.647	51	1.342	50	0.000	35	68.966	70	4.167	51	
		F2U2	15.000	17	13.333	18	20.588	22	0.671	22	0.000	22	70.690	19	8.333	25	
		F2U3	15.000	11	13.333	14	11.765	14	0.671	15	1.190	16	77.586	14	4.167	15	
		F2U4	15.000	11	13.333	12	14.706	13	0.671	14	1.190	16	77.586	14	4.167	13	
		F2U5	15.000	11	13.333	12	14.706	13	0.671	14	2.381	14	77.586	14	4.167	13	
		F2U6	15.000	11	13.333	12	14.706	13	0.671	14	2.381	14	77.586	14	4.167	13	
F3U1		25.000	36	13.333	33	17.647	51	1.342	50	0.000	35	68.966	70	4.167	51		
F3U2		15.000	17	13.333	18	20.588	22	0.671	22	0.000	22	70.690	19	8.333	25		
F3U3	15.000	11	13.333	14	11.765	14	0.671	15	1.190	16	77.586	14	4.167	15			
F3U4	15.000	11	13.333	12	14.706	13	0.671	14	1.190	16	77.586	14	4.167	13			
F3U5	15.000	11	13.333	12	14.706	13	0.671	14	2.381	14	77.586	14	4.167	13			
F3U6	15.000	11	13.333	12	14.706	13	0.671	14	2.381	14	77.586	14	4.167	13			
CFL3	20.000	19	6.667	20	5.882	27	0.671	23	1.190	20	70.690	32	4.167	24			

**Table 4.** Fitness measure  $F$  results for microarray datasets.

		Colon	DLBCL	Leukemia	Lung	Ovarian	11 Tumors	Leukemia 2		
FIXED TH	DESIGN CT	100%	27.0000	13.6670	27.7840	3.4653	25.7300	57.8050	36.4450	
		50%	26.8330	9.0553	27.6170	1.0615	6.5159	56.4890	16.8330	
		25%	26.7500	8.9720	27.5340	0.5307	4.8453	46.0610	16.7500	
		10%	26.7000	8.9220	27.4840	0.4807	3.2079	46.0110	16.7000	
		log2(n)	13.3350	4.4455	9.8046	0.4478	0.0003	47.1270	8.3337	
	AUTOMATIC THRESHOLD	DESIGN CT	F1MIN	26.7050	4.4455	9.8046	0.4478	0.0003	47.1270	8.3337
			F2MIN	13.3350	4.4455	9.8046	0.4478	0.0003	47.1270	8.3337
			F3MIN	13.3350	4.4455	9.8046	0.4478	0.0003	45.9780	8.3337
		DESIGN TC	F1U1	23.3490	0.0039	11.7670	0.8962	0.0008	45.9800	5.5574
			F1U2	23.3460	4.4469	13.7260	0.4480	0.0005	47.1270	5.5563
F1U3			16.6700	8.8899	7.8440	0.4478	0.7940	51.7240	2.7782	
F1U4			10.0020	8.8897	9.8046	0.4478	0.7940	51.7240	2.7782	
F1U5			10.0020	8.8897	9.8046	0.4478	1.5876	51.7240	2.7782	
F1U6			10.0020	8.8897	9.8046	0.4478	1.5876	51.7240	2.7782	
F2U1			16.6730	8.8914	11.7670	0.8962	0.0008	45.9790	2.7793	
F2U2	10.0030		8.8902	13.7260	0.4480	0.0005	47.1270	5.5563		
F2U3	10.0020		8.8898	7.8440	0.4478	0.7940	51.7240	2.7782		
F2U4	10.0020		8.8897	9.8046	0.4478	0.7940	51.7240	2.7782		
F2U5	10.0020		8.8897	9.8046	0.4478	1.5876	51.7240	2.7782		
F2U6	10.0020		8.8897	9.8046	0.4478	1.5876	51.7240	2.7782		
F3U1	16.6730		8.8914	11.7670	0.8962	0.0008	45.9790	2.7793		
F3U2	10.0030		8.8902	13.7260	0.4480	0.0005	47.1270	5.5563		
F3U3	10.0020		8.8898	7.8440	0.4478	0.7940	51.7240	2.7782		
F3U4	10.0020		8.8897	9.8046	0.4478	0.7940	51.7240	2.7782		
F3U5	10.0020		8.8897	9.8046	0.4478	1.5876	51.7240	2.7782		
F3U6	10.0020	8.8897	9.8046	0.4478	1.5876	51.7240	2.7782			
CFL3	13.3360	4.4461	3.9229	0.4480	0.7941	47.1280	2.7785			