

# 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	
FIX 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
AUTOMATIC THRESHOLD	DESIGN CT	MINF1	25.125	9	32.333	9	33.398	7	59.458	14	19.219	21
		MINF2	26.190	6	32.333	9	33.188	5	67.667	9	33.824	8
		MINF3	26.190	6	32.333	9	33.188	5	67.667	9	33.824	8
		MINCF	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
	DESIGN TC	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		18.496	12	31.792	11	31.251	7	60.189	18	20.972	22

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

		Spambase	Madelon	Connect4	Isolet	USPS			
AUTOMATIC THRESHOLD	FIX TH	100%	0.32587	0.53646	0.45142	0.37850	0.20780		
		50%	0.24381	0.37882	0.36797	0.28921	0.13093		
		25%	0.20956	0.32535	0.32453	0.28753	0.10396		
		10%	0.23579	0.30139	0.29932	0.34711	0.15011		
	DESIGN TC	DESIGN CT	MINF1	0.23569	0.27244	0.30609	0.49927	0.17842	
			MINF2	0.23579	0.27244	0.29641	0.56632	0.28708	
			MINF3	0.23579	0.27244	0.29641	0.56632	0.28708	
			MINCF	0.23579	0.27244	0.29641	0.56632	0.28708	
		DESIGN TC	F1U	F1U1	0.25559	0.29060	0.34606	0.38236	0.09543
				F1U2	0.23916	0.26860	0.29139	0.49003	0.17786
				F1U3	0.23587	0.26860	0.28820	0.55217	0.23136
				F1U4	0.23659	0.26827	0.28550	0.57091	0.26441
				F1U5	0.23678	0.26932	0.28461	0.57091	0.26441
				F1U6	0.23678	0.26932	0.28461	0.57091	0.26441
			F2U	F2U1	0.21440	0.29060	0.34606	0.38236	0.10982
				F2U2	0.19393	0.26860	0.29139	0.49003	0.22011
				F2U3	0.22594	0.26860	0.28820	0.55217	0.26441
				F2U4	0.23642	0.26827	0.28550	0.57091	0.26441
				F2U5	0.26495	0.26932	0.28461	0.57091	0.26441
				F2U6	0.26495	0.26932	0.28461	0.57091	0.26441
			F3U	F3U1	0.21440	0.29060	0.34606	0.38236	0.10982
				F3U2	0.19393	0.26860	0.29139	0.49003	0.22011
				F3U3	0.22594	0.26860	0.28820	0.55217	0.26441
				F3U4	0.23642	0.26827	0.28550	0.57091	0.26441
				F3U5	0.26495	0.26932	0.28461	0.57091	0.26441
				F3U6	0.26495	0.26932	0.28461	0.57091	0.26441
			CFL3	0.18922	0.26860	0.28820	0.50644	0.18909	

**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	FIX 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
	DESIGN TC	MINF1	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	68.966	28	12.500	13	
		MINF2	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13	
		MINF3	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13	
		MINCF	20.000	11	6.667	12	14.706	13	0.671	14	0.000	14	70.690	14	12.500	13	
		F1U1	25.000	36	13.333	33	17.647	51	1.342	50	0.000	35	68.966	84	8.333	64	
		F1U2	15.000	17	13.333	18	20.588	22	0.671	22	0.000	22	70.690	19	8.333	25	
		F1U3	15.000	11	13.333	14	11.765	14	0.671	15	1.190	16	77.586	14	4.167	15	
		F1U4	15.000	11	13.333	12	14.706	13	0.671	14	1.190	16	77.586	14	4.167	13	
		F1U5	15.000	11	13.333	12	14.706	13	0.671	14	2.381	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	26	13.333	20	5.882	31	1.342	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		
AUTOMATIC THRESHOLD	FIX TH	100%	0.5000	0.3333	0.5098	0.2058	0.4841	0.8851	0.6181	
		50%	0.4167	0.1944	0.4265	0.0945	0.1627	0.7874	0.2917	
		25%	0.3750	0.1528	0.3848	0.0473	0.1012	0.6164	0.2500	
		10%	0.3500	0.1278	0.3598	0.0223	0.0563	0.5914	0.2250	
	DESIGN CT	MINF1	0.1676	0.0561	0.1229	0.0058	0.0002	0.5751	0.1044	
		MINF2	0.1676	0.0561	0.1229	0.0058	0.0002	0.5893	0.1044	
		MINF3	0.1676	0.0561	0.1229	0.0058	0.0002	0.5893	0.1044	
		MINCF	0.1676	0.0561	0.1229	0.0058	0.0002	0.5893	0.1044	
		DESIGN TC	F1U1	0.2113	0.1125	0.1483	0.0119	0.0004	0.5758	0.0704
			F1U2	0.1264	0.1119	0.1721	0.0059	0.0002	0.5893	0.0698
			F1U3	0.1259	0.1117	0.0984	0.0058	0.0101	0.6467	0.0349
			F1U4	0.1259	0.1116	0.1229	0.0058	0.0101	0.6467	0.0349
			F1U5	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349
			F1U6	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349
			F2U1	0.2113	0.1125	0.1483	0.0119	0.0004	0.5756	0.0355
			F2U2	0.1264	0.1119	0.1721	0.0059	0.0002	0.5893	0.0698
			F2U3	0.1259	0.1117	0.0984	0.0058	0.0101	0.6467	0.0349
			F2U4	0.1259	0.1116	0.1229	0.0058	0.0101	0.6467	0.0349
			F2U5	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349
			F2U6	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349
		F3U1	0.2113	0.1125	0.1483	0.0119	0.0004	0.5756	0.0355	
		F3U2	0.1264	0.1119	0.1721	0.0059	0.0002	0.5893	0.0698	
		F3U3	0.1259	0.1117	0.0984	0.0058	0.0101	0.6467	0.0349	
		F3U4	0.1259	0.1116	0.1229	0.0058	0.0101	0.6467	0.0349	
	F3U5	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349		
	F3U6	0.1259	0.1116	0.1229	0.0058	0.0200	0.6467	0.0349		
	CFL3	0.1688	0.1119	0.0497	0.0115	0.0101	0.5895	0.0351		