

# Supplementary material for "Ensembles of Feature Selectors for Class-Imbalanced Datasets"

Aida de Haro-García<sup>a</sup>, Gonzalo Cerruela-García<sup>a</sup>, Nicolás García-Pedrajas<sup>a,\*</sup>

<sup>a</sup>University of Córdoba, Edificio Einstein, Campus Universitario de Rabanales, 14071 Córdoba, Spain

## Tables

Table 1: Summary of datasets characteristics.

	Dataset	Instances	Features	#Minority	#Majority	Ratio
1	abalone19	4,177	10	32	4145	1:129.5
2	abalone9-18	731	10	42	689	1:16.4
3	ads	3,279	1,558	459	2820	1:6.1
4	adult	48,842	105	11687	37155	1:3.2
5	AID1284Morered	362	914	57	305	1:5.4
6	AID1284red	362	103	57	305	1:5.4
7	AID1608Morered	1,033	914	68	965	1:14.2
8	AID1608red	1,033	154	68	965	1:14.2
9	AID362red	4,279	144	60	4219	1:70.3
10	AID373AID439red	59,795	154	13	59782	1:4598.6
11	AID373red	59,788	155	62	59726	1:963.3
12	AID439Morered	69	914	13	56	1:4.3
13	AID439red	69	81	13	56	1:4.3
14	AID456red	9,982	153	27	9955	1:368.7
15	AID604	95,662	154	340	95322	1:280.4
16	AID604AID644_AllRed	59,782	154	67	59715	1:891.3
17	AID644Morered	206	914	67	139	1:2.1
18	AID644red	206	100	67	139	1:2.1
19	AID687AID721red	33,067	153	21	33046	1:1573.6
20	AID687red	33,067	153	94	32973	1:350.8
21	AID688red	27,189	153	248	26941	1:108.6
22	AID721morered	94	914	21	73	1:3.5
23	AID721red	94	87	21	73	1:3.5
24	AID746AID1284red	59,784	154	57	59727	1:1047.8
25	AID746red	59,788	154	366	59422	1:162.4
26	all-aml	72	7,129	25	47	1:1.9
27	antivirus	372	531	72	300	1:4.2
28	arabidopsis	3,396	4,016	27	3369	1:124.8
29	bank	41,188	62	4640	36548	1:7.9
30	barleyblumeria	4,340	64	1570	2770	1:1.8
31	biodeg	1,055	41	356	699	1:2.0
32	block	574,913	11	2093	572820	1:273.7
33	breast-cancer	286	15	85	201	1:2.4
34	cancer	699	9	241	458	1:1.9

\*This work was partly supported by Grant TIN2015-66108-P of the Ministry of Science and Innovation of the Spanish Government.

\*Corresponding author

*Email addresses:* [adeharo@uco.es](mailto:adeharo@uco.es) (Aida de Haro-García), [gcerruela@uco.es](mailto:gcerruela@uco.es) (Gonzalo Cerruela-García), [npedrajas@uco.es](mailto:npedrajas@uco.es) (Nicolás García-Pedrajas)

Table 1: Summary of datasets characteristics (continuation).

	Dataset	Instances	Features	#Minority	#Majority	Ratio
35	carG	1,728	16	69	1659	1:24.0
36	ccds	3,646	4,016	140	3506	1:25.0
37	census	29,926	409	1912	28014	1:14.7
38	central-nervous-system	60	7,129	21	39	1:1.9
39	cervical-B	858	35	55	803	1:14.6
40	cervical-C	858	35	44	814	1:18.5
41	cervical-H	858	35	35	823	1:23.5
42	cervical-S	858	35	74	784	1:10.6
43	chrom19	170,051	5	177	169874	1:959.7
44	chrom21	126,770	1,612	33	126737	1:3840.5
45	clave	10,800	16	1785	9015	1:5.1
46	climate	540	20	46	494	1:10.7
47	colon-tumor	62	2,000	22	40	1:1.8
48	colposcopies-4	287	62	75	212	1:2.8
49	colposcopies-5	287	62	48	239	1:5.0
50	colposcopies-c	287	62	71	216	1:3.0
51	column-2c	310	6	100	210	1:2.1
52	covtypeCD	20,114	54	2747	17367	1:6.3
53	covtypeT	38,501	54	2747	35754	1:13.0
54	default	30,000	23	6636	23364	1:3.5
55	dlbcl-harvard-tumor	77	7,129	19	58	1:3.1
56	dna	50,000	800	120	49880	1:415.7
57	ecoliCP-IM	220	7	77	143	1:1.9
58	ecoliM	336	7	77	259	1:3.4
59	ecoliMU	336	7	35	301	1:8.6
60	ecoliOM	336	7	20	316	1:15.8
61	euthyroid	3,163	44	293	2870	1:9.8
62	fertility	100	9	12	88	1:7.3
63	german	1,000	61	300	700	1:2.3
64	glassBWFP	214	9	70	144	1:2.1
65	glassBWNFP	214	9	76	138	1:1.8
66	glassContainers	214	9	13	201	1:15.5
67	glassNW	214	9	51	163	1:3.2
68	glassTableware	214	9	9	205	1:22.8
69	glassVWFP	214	9	17	197	1:11.6
70	haberman	306	3	81	150	1:1.9
71	hepatitis	2155	19	32	123	1:3.8
72	hERG-12inputs	565	121	13	52	1:4.0
73	hERG-64inputs	665	64	13	52	1:4.0
74	hERG-9inputs	5130	9	26	104	1:4.0
75	hiv1-proteasa	85,938	160	1010	4928	1:4.9
76	hiva	24,229	1,617	149	4080	1:27.4
77	htru2	517,898	8	1639	16259	1:9.9
78	hypothyroidT	63,772	29	291	3481	1:12.0
79	ilpd	5583	10	167	416	1:2.5
80	ionosphere	8351	34	126	225	1:1.8
81	kddcup98	6954	23,549	49	905	1:18.5
82	kidney-disease	2400	44	150	250	1:1.7
83	labor	957	29	20	37	1:1.9
84	lung-cancer-harvard2	3181	12,533	31	150	1:4.8
85	lung-cancer-michigan	196	7,129	10	86	1:8.6
86	lung-cancer-ontario	239	2,880	15	24	1:1.6
87	monks-2	1601	6	206	395	1:1.9
88	musk	16,598	166	1017	5581	1:5.5
89	new-thyroidT	9215	5	35	180	1:5.1
90	nomao	934,465	172	9844	24621	1:2.5
91	opn310k	110,481	200	922	9559	1:10.4
92	optdigitsZ	15,620	64	554	5066	1:9.1
93	ovarian-cancer	5253	15,154	91	162	1:1.8
94	ozone1hr	22,536	72	73	2463	1:33.7
95	ozone8hr	22,534	72	160	2374	1:14.8
96	p53-mutants	23,142	5,408	15	3127	1:208.5
97	parkinsons	5195	22	48	147	1:3.1
98	particle	113,007	50	3650	9357	1:2.6
99	phoneme	15,404	5	1586	3818	1:2.4

Table 1: Summary of datasets characteristics (continuation).

	Dataset	Instances	Features	#Minority	#Majority	Ratio
100	pima	3768	8	268	500	1:1.9
101	polish-1st	17,027	64	271	6756	1:24.9
102	polish-2nd	210,173	64	400	9773	1:24.4
103	polish-3rd	510,503	64	495	10008	1:20.2
104	polish-4th	79,792	64	515	9277	1:18.0
105	polish-5th	95,910	64	410	5500	1:13.4
106	polya	39,255	169	3309	5946	1:1.8
107	profb	1672	67	224	448	1:2.0
108	satimageF	16,435	36	703	5732	1:8.2
109	satimageT	56,435	36	626	5809	1:9.3
110	secom	11,567	590	104	1463	1:14.1
111	segmentO	12,310	19	330	1980	1:6.0
112	seismic	82,583	22	170	2413	1:14.2
113	semeion	61,593	265	161	1432	1:8.9
114	sens-1-gram	23,801	15	502	3299	1:6.6
115	sens-2-gram	53,801	209	502	3299	1:6.6
116	sick	53,772	33	231	3541	1:15.3
117	spambase	34,601	57	1813	2788	1:1.5
118	spect	1267	22	55	212	1:3.9
119	spectf	2349	44	95	254	1:2.7
120	splice-EI	63,175	120	762	2413	1:3.2
121	splice-IE	43,175	120	765	2410	1:3.2
122	steel-b	21,941	27	402	1539	1:3.8
123	steel-d	31,941	27	55	1886	1:34.3
124	steel-k	51,941	27	391	1550	1:4.0
125	steel-p	61,941	27	158	1783	1:11.3
126	steel-s	61,941	27	72	1869	1:26.0
127	steel-z	11,941	27	190	1751	1:9.2
128	thoracic	5470	27	70	400	1:5.7
129	tic-tac-toe	9958	9	332	626	1:1.9
130	tis	11,337	927	331	1006	1:3.0
131	titanic	22,201	8	711	1490	1:2.1
132	ustilago	46,141	4,016	65	6076	1:93.5
133	vehicleVAN	3846	18	199	647	1:3.3
134	voice	3126	310	42	84	1:2.0
135	vowelZ	2990	10	90	900	1:10.0
136	wilt	94,839	5	261	4578	1:17.5
137	yeastCYT-POX	8483	8	20	463	1:23.2
138	yeastEXC	61,484	8	35	1449	1:41.4
139	yeastME1	11,484	8	44	1440	1:32.7
140	yeastME2	11,484	8	51	1433	1:28.1