

Table S1: Details of the statistical parameters calculated for the constructed models

Construct ed total Models	Dataset	Activity Threshold	No. of Compounds	Tranin g	Tes t	Confusion Matrix						Sensitivi ty	Specifici ty	RO C	G-me an	F-measu re	MC C	Accura cy	
						T P	F N	T N	F P	Tota l P	Total N								Tot al
Model 1	Dataset-1	< 10 and >10	70	45	25	0	2	23	0	23	2	25	1.00	0.00	0.52	0.00	0.96	0.00	0.92
Model 2		< 30 and >30	70	45	25	6	1	16	2	18	7	25	0.89	0.86	0.86	0.87	0.91	0.72	0.88
Model 3		< 50 and >50	70	45	25	10	0	14	1	15	10	25	0.93	1.00	1.00	0.97	0.96	0.92	0.96
Model 4		< 100 and >100	70	45	25	13	0	11	1	12	13	25	0.917	1.00	0.93	0.96	0.96	0.92	0.96
Model 5	Dataset-2	< 10 and >10	39	25	14	0	1	13	0	13	1	14	1.00	0.00	0.54	0.00	0.96	0.00	0.93
Model 6		< 30 and >30	39	25	14	10	2	1	1	2	12	14	0.50	0.83	0.91	0.64	0.40	0.28	0.78
Model 7		< 50 and >50	39	25	14	5	0	8	1	9	5	14	0.89	1.00	0.88	0.94	0.94	0.86	0.93
Model 8		< 100 and >100	39	25	14	0	0	14	0	14	0	14	1.00	0.00	non e	0.00	1.00	0.00	1.00
Model 9	Dataset-3	< 10 and >10	65	42	23	0	0	23	0	23	0	23	1.00	0.00	non e	0.00	1.00	0.00	1.00
Model 10		< 30 and >30	65	42	23	1	0	22	0	22	1	23	1.00	1.00	0.52	1.00	1.00	1.00	1.00
Model 11		< 50 and >50	65	42	23	1	1	19	2	21	2	23	0.90	0.50	0.92	0.67	0.93	0.34	0.87
Model 12		< 100 and >100	65	42	23	6	0	17	0	17	6	23	1.00	1.00	0.92	1.00	1.00	1.00	1.00
Model 13	Dataset-4	< 10 and >10	85	45	40	1	0	19	5	24	1	25	0.79	1.00	0.50	0.89	0.88	0.36	0.80
Model 14		< 30 and >30	85	45	40	4	1	19	1	20	5	25	0.95	0.80	0.52	0.87	0.95	0.75	0.92
Model 15		< 50 and >50	85	45	40	10	0	10	5	15	10	25	0.67	1.00	0.92	0.82	0.80	0.67	0.80

Model 16		< 100 and >100	85	45	40	1 7	0	1	7	8	17	25	0.12	1.00	0.6 2	0.35	0.22	0.30	0.72
Model 17	Dataset-5	< 10 and >10	64	41	23	1	0	20	2	22	1	23	0.91	1.00	0.5 2	0.95	0.95	0.55	0.91
Model 18		< 30 and >30	64	41	23	4	0	18	1	19	4	23	0.95	1.00	0.8 8	0.97	0.97	0.87	0.96
Model 19		< 50 and >50	64	41	23	6	0	16	1	17	6	23	0.94	1.00	0.8 3	0.97	0.97	0.90	0.96
Model 20		< 100 and >100	64	41	23	9	0	8	6	14	9	23	0.57	1.00	0.8 2	0.75	0.77	0.59	0.74
Model 21	Dataset-6	< 10 and >10	70	45	25	1	1	23	0	23	2	25	1.00	0.5	0.5 2	0.71	0.98	0.69	0.96
Model 22		< 30 and >30	70	45	25	7	0	16	2	18	7	25	0.89	1.00	0.9 1	0.94	0.94	0.83	0.92
Model 23		< 50 and >50	70	45	25	6	4	15	0	15	10	25	1.00	0.60	0.8 8	0.77	0.88	0.69	0.84
Model 24		< 100 and >100	70	45	25	1 2	1	11	1	12	13	25	0.92	0.92	0.9 6	0.92	0.92	0.84	0.92
Model 25	combined	< 10 and >10	309	200	10 9	5	3	10 1	0	101	8	109	1.00	0.62	0.7 4	0.79	0.98	0.78	0.97
Model 26		< 30 and >30	309	200	10 9	1 2	10	85	2	87	22	109	0.98	0.54	0.7 9	0.73	0.93	0.63	0.89
Model 27		< 50 and >50	309	200	10 9	2 8	8	67	6	73	36	109	0.92	0.78	0.8 6	0.84	0.90	0.71	0.87

Figure S1: Overall KNIME workflow of classification models.

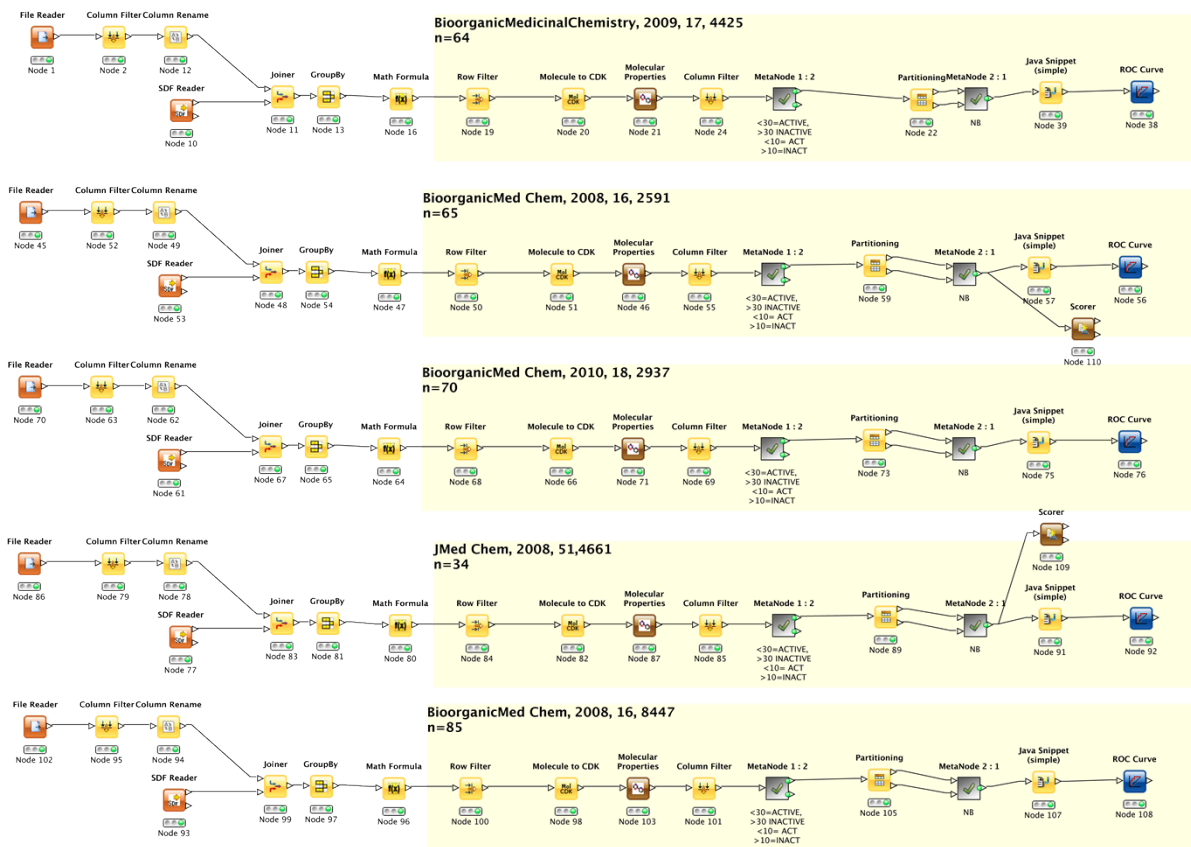


Figure S2: Loading scatter plot for the descriptors

