Experiment results of Explanations as a new metric for feature selection: a systematic approach

Haomiao Wang, Emmanuel Doumard, Chantal Soulé-Dupuy, Philippe Kémoun, Julien Aligon †,*, Paul Monsarrat †

I. Dataset Filters determination

A. Number of Features

A sufficient number of features is necessary for a meaningful feature selection in the context of XAI, 10 was chosen as the lower limit. The upper limit was set to 150, which ensures acceptable computation time and covers most of the OpenML datasets meeting the conditions (binary classification tasks, continuous explanatory features only, no missing data).

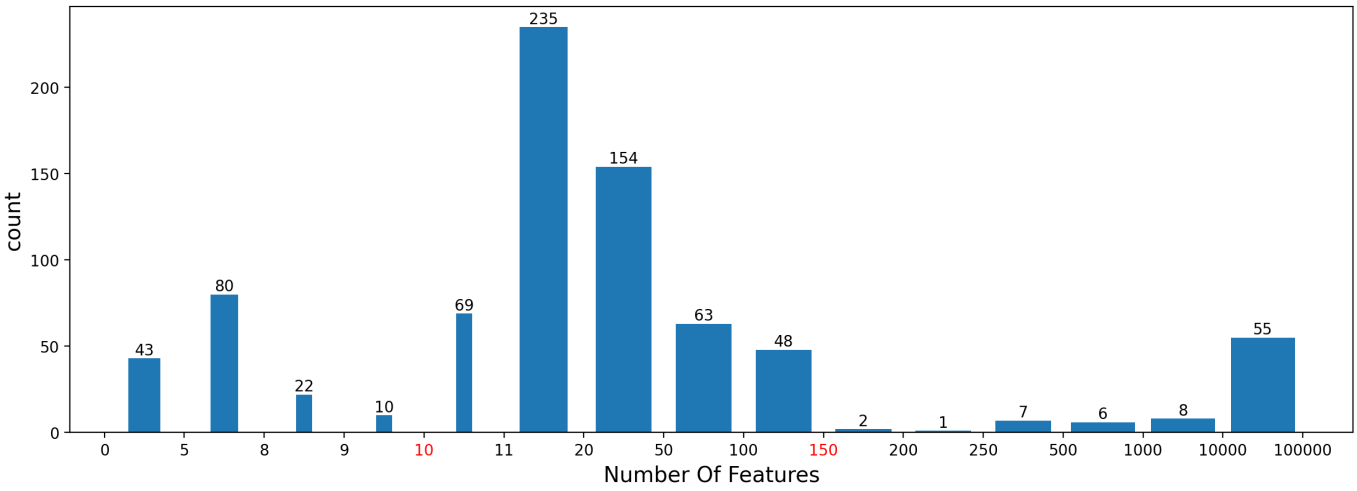


Fig. 1: **Frequency histogram of the number of features.** The x-axis scale is non-linear, the bar width is proportional to the interval size.

B. Number of Instances

As shown in Figure 2, the majority of the datasets which met the conditions (binary classification tasks, continuous explanatory features only, no missing data) on OpenML contain 5,000 instances or less. In this experiment, the computation time of explanation for a dataset containing more than 10,000 instances was overlong. Since there is a trough between 12,000 and 13,000 instances, 12,000 was chosen as the cut-off.

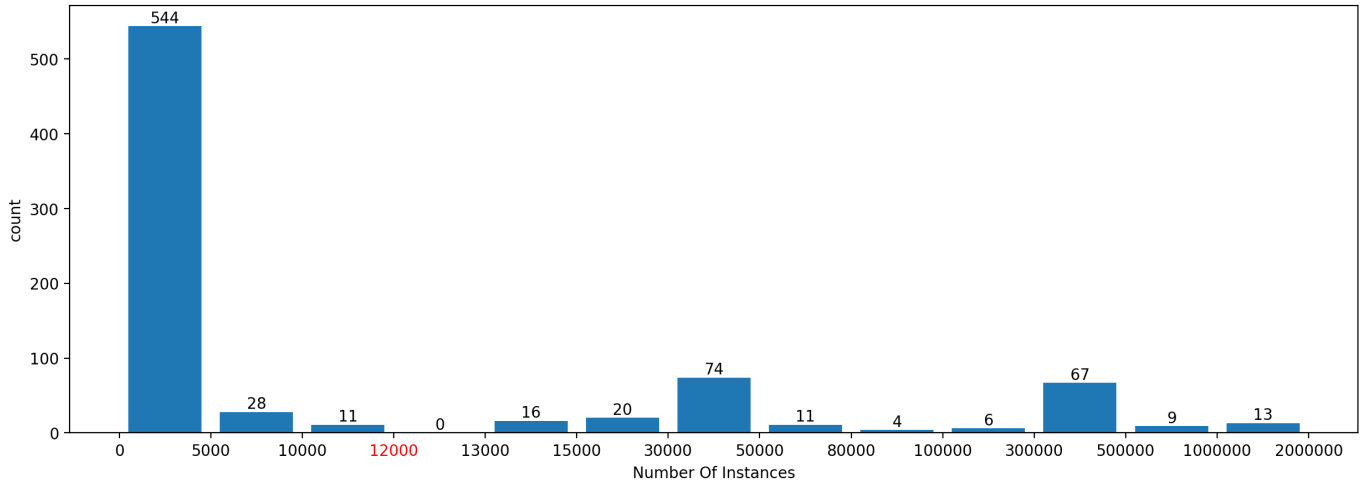


Fig. 2: **Frequency histogram of the number of instances.** The x-axis scale is non-linear, the bar width is proportional to the interval size.

C. Datasets information

TABLE I: Datasets information

id	name	#OfClasses	# OfFeatures	# OfInstances	MajorityClassSize
40	sonar	2	61	208	111
44	spambase	2	58	4601	2788
53	heart-statlog	2	14	270	150
59	ionosphere	2	35	351	225
311	oil_spill	2	50	937	896
715	fri_c3_1000_25	2	26	1000	557
716	fri_c3_100_50	2	51	100	62
717	rmftsa_ladata	2	11	508	286
718	fri_c4_1000_100	2	101	1000	564
721	pwLinear	2	11	200	103
723	fri_c4_1000_25	2	26	1000	547
732	fri_c0_250_50	2	51	250	133
735	cpu_small	2	13	8192	5715
740	fri_c3_1000_10	2	11	1000	560
742	fri_c4_500_100	2	101	500	283
746	fri_c1_250_25	2	26	250	143
752	puma32H	2	33	8192	4128
753	wisconsin	2	33	194	104
756	autoPrice	2	16	159	105
758	analcatdata_election2000	2	15	67	49
759	analcatdata_olympic2000	2	12	66	33
761	cpu_act	2	22	8192	5715
762	fri_c2_100_10	2	11	100	55
763	fri_c0_250_10	2	11	250	125
766	fri_c1_500_50	2	51	500	262
768	fri_c3_100_25	2	26	100	55
769	fri_c1_250_50	2	51	250	137
773	fri_c0_250_25	2	26	250	126
775	fri_c2_100_25	2	26	100	57
778	bodyfat	2	15	252	128
779	fri_c1_500_25	2	26	500	267
783	fri_c3_100_10	2	11	100	60
785	wind_correlations	2	47	45	23

TABLE I: (continued from previous page)

id	name	#OfClasses	#OfFeatures	#OfInstances	MajorityClassSize
788	triazines	2	61	186	109
789	fri_c1_100_10	2	11	100	53
793	fri_c3_250_10	2	11	250	135
794	fri_c2_250_25	2	26	250	139
797	fri_c4_1000_50	2	51	1000	560
800	pyrim	2	28	74	43
805	fri_c4_500_50	2	51	500	264
806	fri_c3_1000_50	2	51	1000	555
812	fri_c1_100_25	2	26	100	53
820	chatfield_4	2	13	235	142
824	fri_c1_500_10	2	11	500	274
828	fri_c4_100_100	2	101	100	53
830	fri_c2_250_10	2	11	250	159
833	bank32nh	2	33	8192	5649
834	fri_c4_250_100	2	101	250	140
837	fri_c1_1000_50	2	51	1000	547
838	fri_c4_500_25	2	26	500	284
845	fri_c0_1000_10	2	11	1000	509
847	wind	2	15	6574	3501
849	fri_c0_1000_25	2	26	1000	503
850	fri_c0_100_50	2	51	100	51
851	tecator	2	125	240	138
855	fri_c4_500_10	2	11	500	276
863	fri_c4_250_10	2	11	250	133
866	fri_c2_1000_50	2	51	1000	582
868	fri_c4_100_25	2	26	100	54
869	fri_c2_500_10	2	11	500	286
873	fri_c3_250_50	2	51	250	142
876	fri_c1_100_50	2	51	100	56
879	fri_c2_500_25	2	26	500	304
880	mu284	2	11	284	142
882	pollution	2	16	60	31
888	fri_c0_500_50	2	51	500	256
889	fri_c0_100_25	2	26	100	50
896	fri_c3_500_25	2	26	500	280
903	fri_c2_1000_25	2	26	1000	563
904	fri_c0_1000_50	2	51	1000	510
910	fri_c1_1000_10	2	11	1000	564
913	fri_c2_1000_10	2	11	1000	580
917	fri_c1_1000_25	2	26	1000	546
918	fri_c4_250_50	2	51	250	135
920	fri_c2_500_50	2	51	500	295
922	fri_c2_100_50	2	51	100	58
926	fri_c0_500_25	2	26	500	255
927	hutsof99_child_witness	2	16	42	25
933	fri_c4_250_25	2	26	250	136
935	fri_c1_250_10	2	11	250	140
936	fri_c3_500_10	2	11	500	272
937	fri_c3_500_50	2	51	500	282
943	fri_c0_500_10	2	11	500	259
958	segment	2	20	2310	1980
970	analcatdata_authorship	2	71	841	524
971	mfeat-fourier	2	77	2000	1800
973	wine	2	14	178	107
976	JapaneseVowels	2	15	9961	8347

TABLE I: (continued from previous page)

id	name	#OfClasses	#OfFeatures	#OfInstances	MajorityClassSize
979	waveform-5000	2	41	5000	3308
980	optdigits	2	65	5620	5048
994	vehicle	2	19	846	628
995	mfeat-zernike	2	48	2000	1800
1004	synthetic_control	2	61	600	500
1019	pendigits	2	17	10992	9848
1020	mfeat-karhunen	2	65	2000	1800
1021	page-blocks	2	11	5473	4913
1045	kc1-top5	2	95	145	137
1049	pc4	2	38	1458	1280
1050	pc3	2	38	1563	1403
1054	mc2	2	40	161	109
1056	mc1	2	39	9466	9398
1059	ar1	2	30	121	112
1060	ar3	2	30	63	55
1061	ar4	2	30	107	87
1062	ar5	2	30	36	28
1063	kc2	2	22	522	415
1064	ar6	2	30	101	86
1065	kc3	2	40	458	415
1066	kc1-binary	2	95	145	85
1067	kc1	2	22	2109	1783
1068	pc1	2	22	1109	1032
1069	pc2	2	37	5589	5566
1071	mw1	2	38	403	372
1441	KungChi3	2	40	123	107
1442	MegaWatt1	2	38	253	226
1443	PizzaCutter1	2	38	661	609
1444	PizzaCutter3	2	38	1043	916
1446	CostaMadre1	2	38	296	258
1447	CastMetal1	2	38	327	285
1448	KnuggetChase3	2	40	194	158
1450	MindCave2	2	40	125	81
1451	PieChart1	2	38	705	644
1452	PieChart2	2	37	745	729
1453	PieChart3	2	38	1077	943
1487	ozone-level-8hr	2	73	2534	2374
1488	parkinsons	2	23	195	147
1490	planning-relax	2	13	182	130
1494	qsar-biodeg	2	42	1055	699
1496	ringnorm	2	21	7400	3736
1504	steel-plates-fault	2	34	1941	1268
1507	twonorm	2	21	7400	3703
1510	wdbc	2	31	569	357
1547	autoUniv-au1-1000	2	21	1000	741
1566	hill-valley	2	101	1212	612
1600	SPECTF	2	45	267	212
40900	Satellite	2	37	5100	5025
40994	climate-model-simulation-crashes	2	21	540	494
41146	sylvine	2	21	5124	2562
41156	ada	2	49	4147	3118
41730	FOREX_usdchf-day-High	2	12	1835	919
41872	FOREX_eurhkd-day-Close	2	12	1832	917
41897	FOREX_usddkk-day-Close	2	12	1832	916
41945	ilpd-numeric	2	11	583	416

TABLE I: (continued from previous page)

id	name	#OfClasses	#OfFeatures	#OfInstances	MajorityClassSize
41946	Sick_numeric	2	30	3772	3541
41978	TuningSVMs	2	81	156	94

II. Model tuning

A. RFECV grid search

TABLE II: Hyperparameters of the grid search in RFECV

Hyperparameter	Values
max_depth	$\{3, 5, 8\}$
max_samples	$\{0.5, 0.6, 0.7, 0.8, 0.9\}$

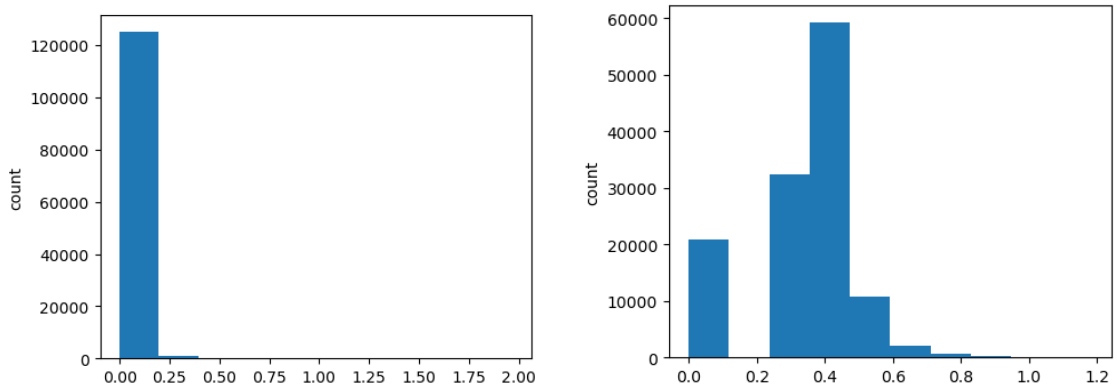
B. Training grid search

TABLE III: Hyperparameters of the grid search in model training

Model	Hyperparameter	Values
en	l1_ratio	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0,9
knn	n_neighbors	3, 4, 5, 6, 7, 8
xg	max_depth	2, 3, 4
	min_child_weight	1, 2, 3, 4,
	gamma	1, 2, 3, 4
	eta	0.01

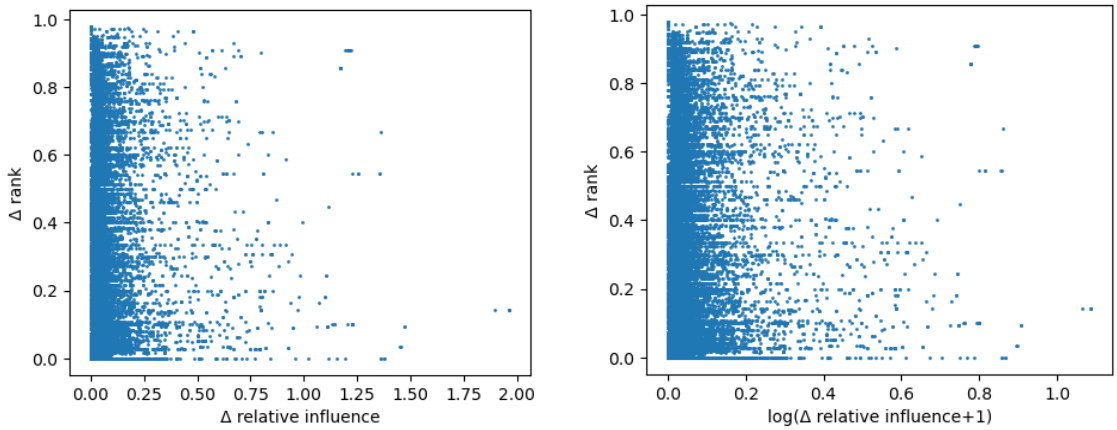
III. RI Metric Construction

Scaling is required for *Relative influence change* due to its distribution (Figure 3(a)). After testing various scaling (Figure 4), the fourth root transformation was applied. This scaling made the distribution of *Relative influence change* more normal (if ignoring 0), and adjusted the two components of the RI metric (rank and influence changes) to a similar scale (Figure(4(d)).

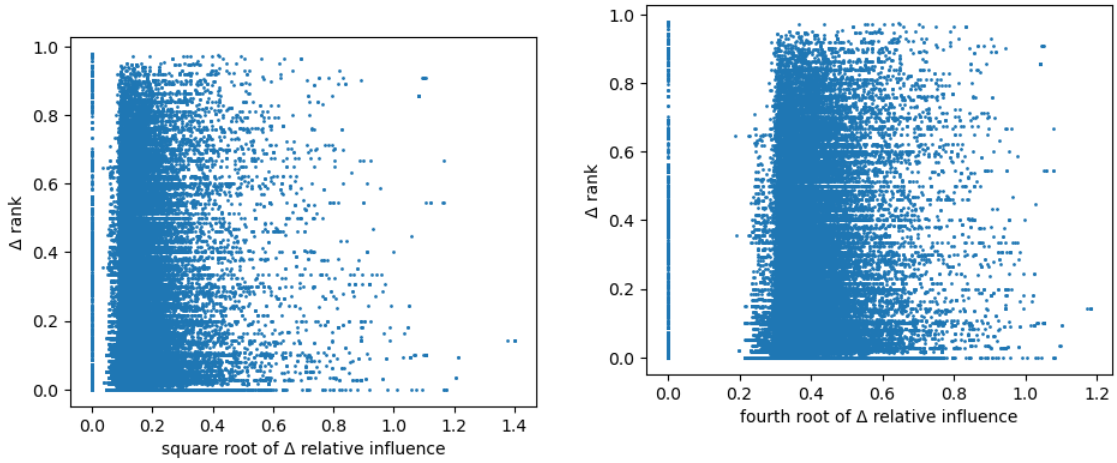


(a) No transformation (b) Fourth root transformation

Fig. 3: **Frequency histogram of Relative influence change**



(a) No transformation (b) Logarithmic transformation



(c) Square root transformation (d) Fourth root transformation

Fig. 4: **Scaling of Relative influence change**

IV. Overall result

A. Statistical analysis

TABLE IV: Statistical results of the *Kendall's* τ

		fisher	reliefF	spec	f	chi2	rfs	mrmr	cmim	jmi	borutashap	rf	svm	lg
en	avg.	0.367	0.475	0.230	0.374	0.333	0.350	0.323	0.400	0.375	0.362	0.403	0.383	0.360
	std.	0.532	0.501	0.317	0.530	0.530	0.436	0.416	0.510	0.474	0.555	0.604	0.436	0.478
knn	avg.	0.365	0.472	0.211	0.370	0.416	0.318	0.397	0.40	0.432	0.351	0.376	0.306	0.360
	std.	0.521	0.488	0.386	0.517	0.482	0.434	0.401	0.516	0.473	0.563	0.527	0.430	0.430
nb	avg.	0.316	0.433	0.294	0.327	0.347	0.278	0.368	0.323	0.328	0.382	0.366	0.362	0.354
	std.	0.564	0.532	0.374	0.554	0.546	0.412	0.453	0.553	0.477	0.545	0.545	0.399	0.499
xg	avg.	0.496	0.494	0.235	0.503	0.555	0.405	0.396	0.518	0.516	0.577	0.556	0.397	0.403
	std.	0.510	0.518	0.344	0.510	0.452	0.401	0.411	0.453	0.450	0.493	0.483	0.373	0.443

TABLE V: Statistical results of the *Relative influence change*

		fisher	reliefF	spec	f	chi2	rfs	mrmr	cmim	jmi	borutashap	rf	svm	lg
en	avg.	0.140	0.114	0.128	0.140	0.146	0.155	0.123	0.123	0.107	0.125	0.145	0.141	0.135
	std.	0.157	0.125	0.144	0.157	0.175	0.203	0.153	0.166	0.126	0.139	0.158	0.165	0.155
knn	avg.	0.132	0.098	0.122	0.131	0.137	0.132	0.117	0.12	0.115	0.129	0.132	0.129	0.124
	std.	0.135	0.085	0.126	0.133	0.150	0.157	0.143	0.153	0.146	0.141	0.137	0.136	0.124
nb	avg.	0.116	0.101	0.120	0.118	0.123	0.136	0.098	0.094	0.081	0.125	0.129	0.153	0.134
	std.	0.130	0.113	0.138	0.134	0.146	0.181	0.135	0.138	0.090	0.159	0.162	0.205	0.170
xg	avg.	0.142	0.117	0.127	0.140	0.142	0.121	0.139	0.131	0.124	0.139	0.151	0.107	0.118
	std.	0.157	0.126	0.156	0.155	0.154	0.139	0.165	0.156	0.141	0.146	0.166	0.099	0.114

TABLE VI: Statistical results of the *RI* metric

		fisher	reliefF	spec	f	chi2	rfs	mrmr	cmim	jmi	borutashap	rf	svm	lg
en	avg.	0.127	0.077	0.146	0.122	0.115	0.099	0.127	0.130	0.117	0.125	0.151	0.096	0.094
	std.	0.176	0.063	0.189	0.169	0.148	0.131	0.134	0.149	0.122	0.130	0.195	0.146	0.138
knn	avg.	0.117	0.069	0.135	0.115	0.106	0.113	0.133	0.123	0.111	0.121	0.130	0.133	0.117
	std.	0.163	0.061	0.140	0.160	0.161	0.104	0.139	0.136	0.123	0.164	0.178	0.133	0.127
nb	avg.	0.085	0.074	0.109	0.085	0.075	0.103	0.102	0.102	0.099	0.099	0.112	0.108	0.092
	std.	0.115	0.070	0.121	0.115	0.109	0.085	0.073	0.089	0.080	0.103	0.122	0.091	0.080
xg	avg.	0.085	0.089	0.110	0.082	0.087	0.099	0.109	0.094	0.082	0.062	0.063	0.102	0.100
	std.	0.111	0.089	0.095	0.101	0.106	0.087	0.097	0.117	0.087	0.069	0.076	0.067	0.083

TABLE VII: Statistical results of the *RIA* metric

		fisher	reliefF	spec	f	chi2	rfs	mrmr	cmim	jmi	borutashap	rf	svm	lg
en	avg.	0.0001	-0.0029	-0.0010	0.0001	-0.0028	0.0019	-0.0039	-0.0030	-0.0027	-0.0027	-0.0038	0.0020	0.0022
	std.	0.0154	0.0119	0.0122	0.0154	0.0191	0.0132	0.0166	0.0137	0.0142	0.0176	0.0177	0.0128	0.0149
knn	avg.	0.0046	0.0028	0.0016	0.0046	0.0044	0.0052	0.0053	0.0071	0.0070	0.0066	0.0076	0.0060	0.0047
	std.	0.0163	0.0067	0.0124	0.0163	0.0219	0.0160	0.0240	0.0219	0.0221	0.0213	0.0226	0.0168	0.0190
nb	avg.	0.0058	-0.0039	-0.0011	0.0058	0.0049	0.0006	-0.0041	-0.0037	-0.0052	-0.0022	-0.0007	0.0002	0.0014
	std.	0.0498	0.0100	0.0253	0.0496	0.0487	0.0157	0.0113	0.0297	0.0152	0.0156	0.0234	0.0182	0.0190
xg	avg.	-0.0011	-0.0008	-0.0013	-0.0011	-0.0011	-0.0012	-0.0014	-0.0015	-0.0010	-0.0003	-0.0005	-0.0010	-0.0014
	std.	0.0043	0.0026	0.0038	0.0043	0.0043	0.0061	0.0029	0.0037	0.0027	0.0021	0.0022	0.0058	0.0049

B. 3D analysis

Positioning the feature selection methods in a tridimensional space with axes that represent respectively explanation (RI), accuracy and retention rate. The following figures demonstrate the 3D analysis in the different models.

In general, the results in the *knn* model (Figure 6) and in the *nb* model (Figure 7) were similar to the *en* model (Figure 5), the optimal FS method was different for each dimension (retention rate, accuracy and explanation). Whereas in the *xg* model (Figure 8), *borutashap* and *rf* were close to the global optimum.

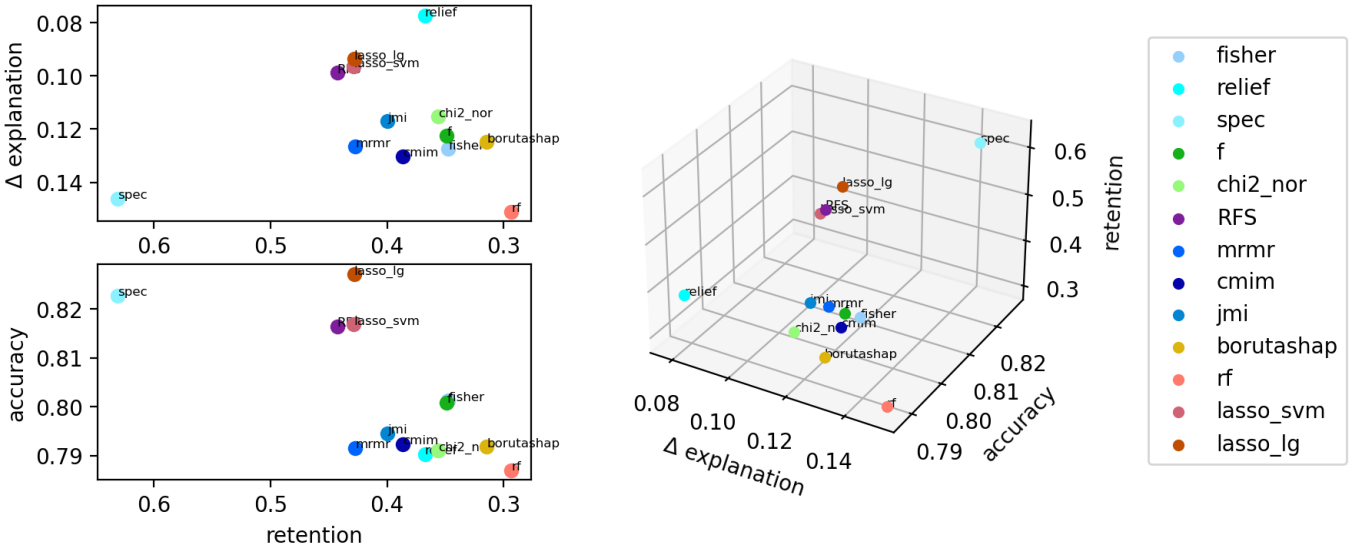


Fig. 5: 3D analysis in the *en* model.

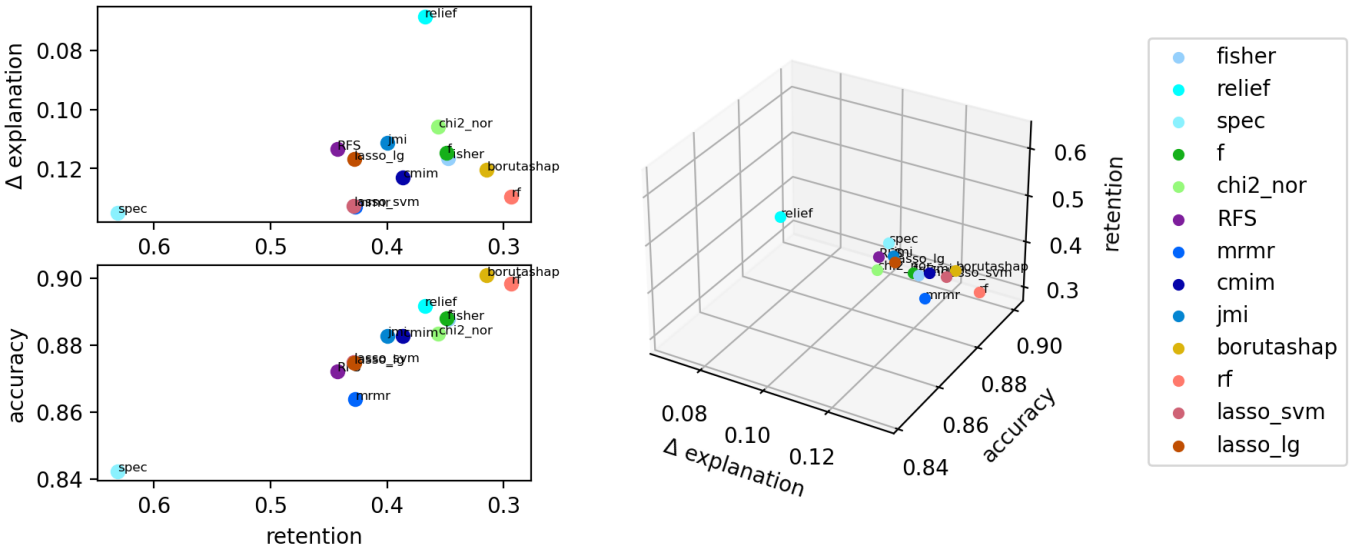


Fig. 6: 3D analysis in the *knn* model.

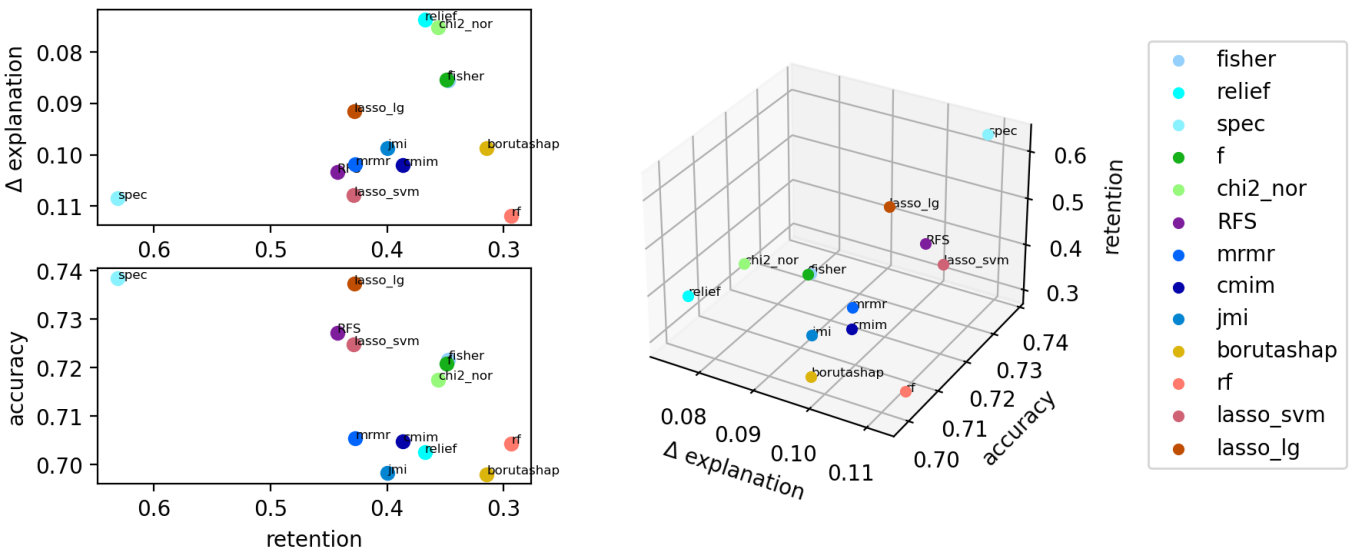


Fig. 7: 3D analysis in the *nb* model.

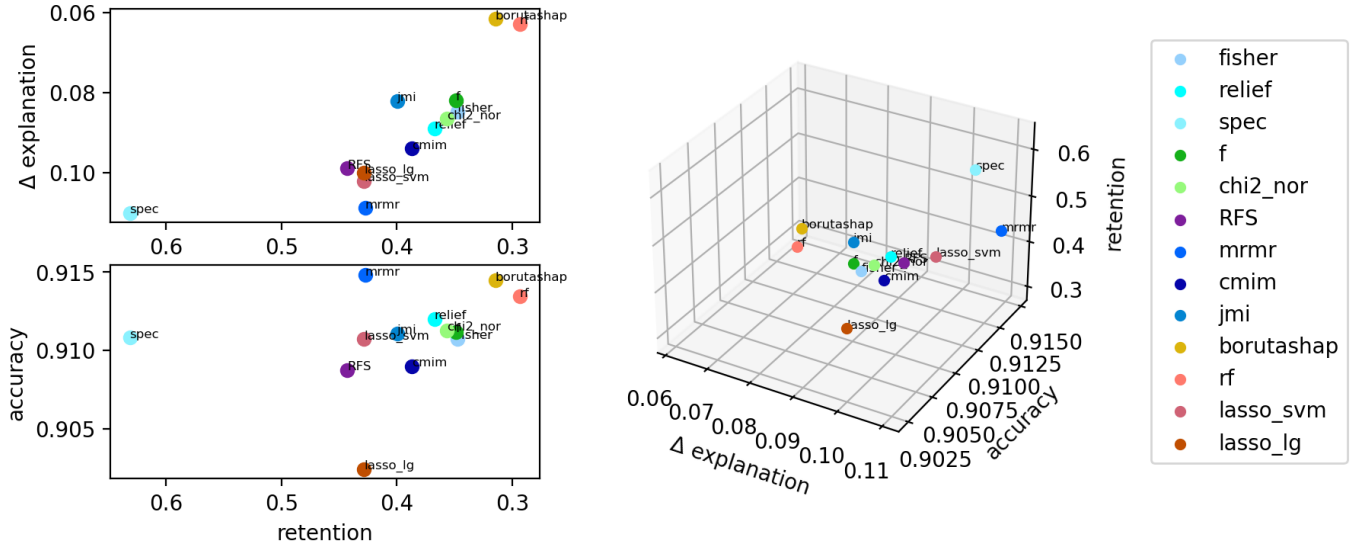


Fig. 8: 3D analysis in the *xg* model.

V. Use cases

A. Metrics

In the Table VIII and IX, *RR* stands for *Retention Rate*, τ for *Kendall rank correlation coefficient* and *rinfc* for *Relative influence change*.

TABLE VIII: Evaluation of Feature Selection Methods of the *Oxford Parkinson's Disease Detection dataset*

FS	RR	ML Models															
		en				knn				nb				xg			
		τ	rinfc	RI	RIA	τ	rinfc	RI	RIA	τ	rinfc	RI	RIA	τ	rinfc	RI	RIA
fisher	9.09%	1	0.4079	0.2339	0.0060	-1	0.4527	0.3045	-0.0094	1	0.1457	0.3057	-0.0282	-1	0.3951	0.0713	0.0004
reliefF	22.73%	0.8	0.1508	0.0315	-0.0026	1	0.0786	0.0421	≈0	0.2	0.1857	0.0949	-0.0088	-0.4	0.2852	0.2187	0.0022
spec	77.27%	0.1176	0.1020	0.1008	0.0010	0.1324	0.1367	0.1684	0.0095	-0.1765	0.1060	0.1061	-0.0098	0.0441	0.0910	0.0442	0.0002
f	9.09%	1	0.4079	0.2339	0.0060	-1	0.4527	0.3045	-0.0094	1	0.1457	0.3057	-0.0282	-1	0.3951	0.0713	0.0004
chi2	22.73%	-0.2	0.2585	0.1232	-0.0107	0.4	0.2471	0.0811	0.0008	1	0.1920	0.0598	-0.0052	0.2	0.2005	0.1807	0.0046
rfs	50.00%	0.4182	0.0337	0.0307	≈0	0.6364	0.0922	0.0963	≈0	0.2	0.1158	0.1261	-0.0045	0.2727	0.1290	0.1325	0.0061
mrmr	4.55%	/	0.0589	0.5901	-0.0514	/	0.5611	0.1818	0.0009	/	/	/	/	/	0.4941	≈0	≈0
cmim	4.55%	/	0.0589	0.5901	-0.0514	/	0.5611	0.1818	0.0009	/	/	/	/	/	0.4941	≈0	≈0
jmi	4.55%	/	0.0589	0.5901	-0.0514	/	0.5611	0.1818	0.0009	/	/	/	/	/	0.4941	≈0	≈0
borutashap	4.55%	/	0.0589	0.5901	-0.0514	/	0.5611	0.1818	0.0009	/	/	/	/	/	0.4941	≈0	≈0
rf	4.55%	/	0.0589	0.5901	-0.0514	/	0.5611	0.1818	0.0009	/	/	/	/	/	0.4941	≈0	≈0
svm	36.36%	0	0.0764	0.0506	-0.0010	0.4286	0.1027	0.1708	0.0035	0.4286	0.1255	0.1784	-0.0165	0.3571	0.2645	0.1751	0.0063
lg	36.36%	0.7857	0.0690	0.0174	-0.0001	0.5	0.0893	0.1339	≈0	0	0.1221	0.1533	-0.0141	-0.2143	0.2116	0.1937	0.0099

TABLE IX: Evaluation of Feature Selection Methods of the *Indian Liver Patient dataset*

FS	RR	ML Models															
		en				knn				nb				xg			
		τ	Rinfc	RI	RIA	τ	Rinfc	RI	RIA	τ	Rinfc	RI	RIA	τ	Rinfc	RI	RIA
fisher	40%	0.6667	0.2820	0.0756	≈0	1	0.1314	0.1269	-0.0061	1	0.1620	0.0905	0.0029	-0.3333	0.0542	0.0505	0.0010
reliefF	30%	1	0.2086	0.0187	≈0	1	0.0920	≈0	≈0	1	0.0948	≈0	≈0	-0.3333	0.1200	0.1625	0.0033
spec	30%	0.3333	0.2884	0.4416	≈0	-0.3333	0.1410	0.3849	-0.0158	1	0.1953	0.4546	0.1068	-0.3333	0.1854	0.4410	-0.0076
f	40%	0.6667	0.2820	0.0756	≈0	1	0.1314	0.1269	-0.0061	1	0.1620	0.0905	0.0029	-0.3333	0.0542	0.0505	0.0010
chi2	50%	1	0.1034	0.0097	≈0	1	0.0673	0.0516	-0.0023	1	0.0297	0.0201	0.0003	0.4	0.0576	0.0669	0.0022
rfs	60%	0.2	0.3344	0.2197	-0.0008	0.0667	0.2695	0.2173	-0.0075	0.2	0.3143	0.2062	0.0325	0.8667	0.1885	0.1828	0.0016
mrmr	10%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
cmim	10%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
jmi	10%	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/
borutashap	50%	1	0.1034	0.0097	≈0	1	0.0673	0.0516	-0.0023	1	0.0297	0.0201	0.0003	0.4	0.0576	0.0669	0.0022
rf	30%	1	0.3101	0.0773	≈0	1	0.0811	0.0408	-0.0009	1	0.2044	0.0385	-0.0004	-1	0.1292	0.1882	0.0036
svm	60%	0.2	0.3344	0.2197	-0.0008	0.0667	0.2695	0.2173	-0.0075	0.2	0.3143	0.2062	0.0325	0.8667	0.1885	0.1828	0.0016
lg	10%	/	0.3921	0.4967	≈0	/	0.3602	0.7990	-0.0617	/	/	/	/	/	0.0818	0.2892	-0.0050

B. Summary plots

In the following plots, the features are sorted in descending order of their contribution to the model. Each feature is renamed according to its original order in the dataset, the feature name, and the order of importance in the explanation for the full feature set ($v[OriginalOrder]\_[FeatureName]\_[ImportanceOrder]$). For each feature, a dot represents an instance of the dataset (red for a high feature value and blue for a low value). A positive value on the x-axis indicates feature contributes positively to the prediction for this instance, and conversely for a negative value.

1) Oxford Parkinson's Disease Detection dataset:

Fig. 9: Summary plots for the Oxford Parkinson's Disease Detection dataset in the *en* model.

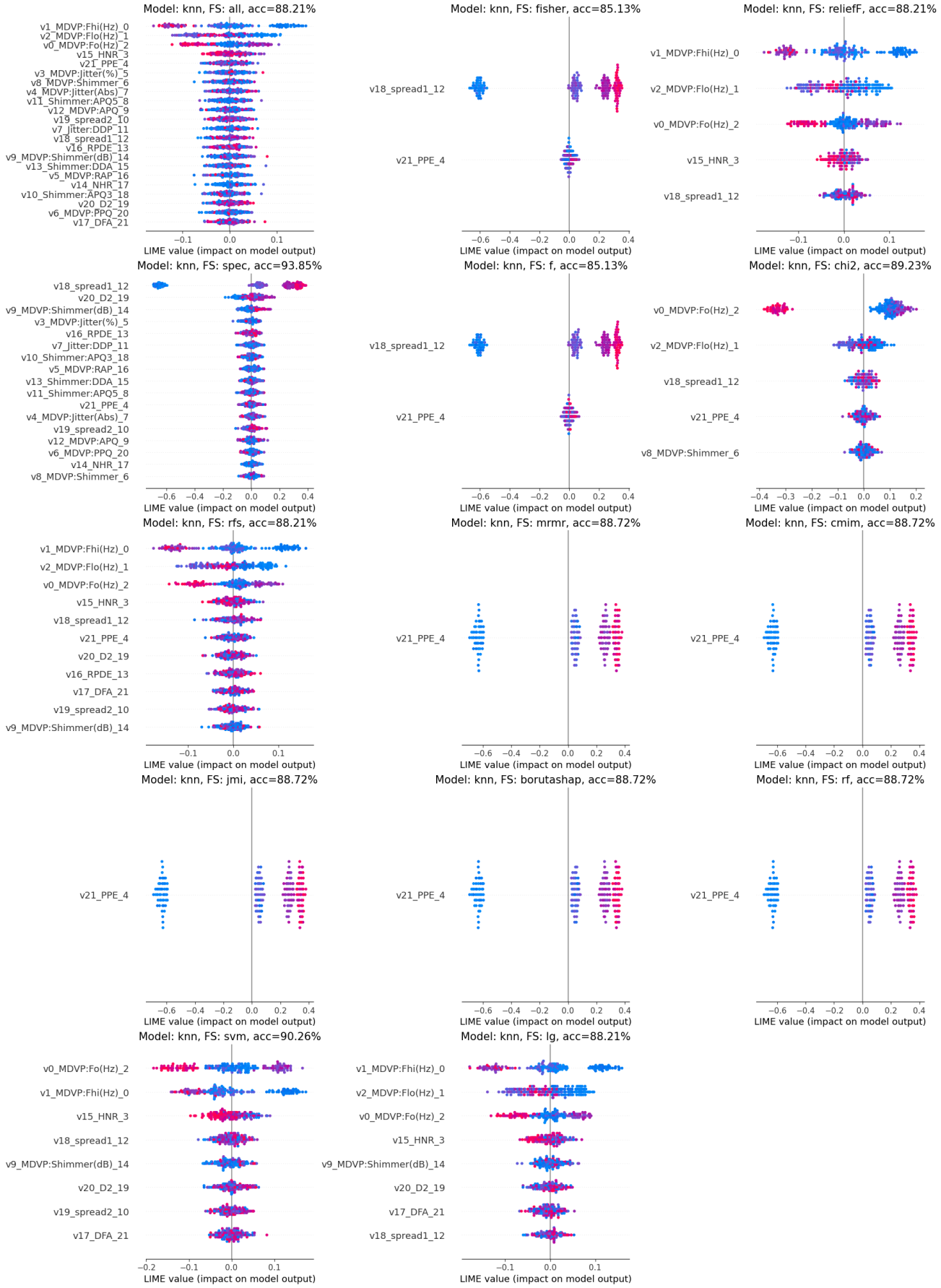


Fig. 10: Summary plots for the Oxford Parkinson's Disease Detection dataset in the *knn* model.

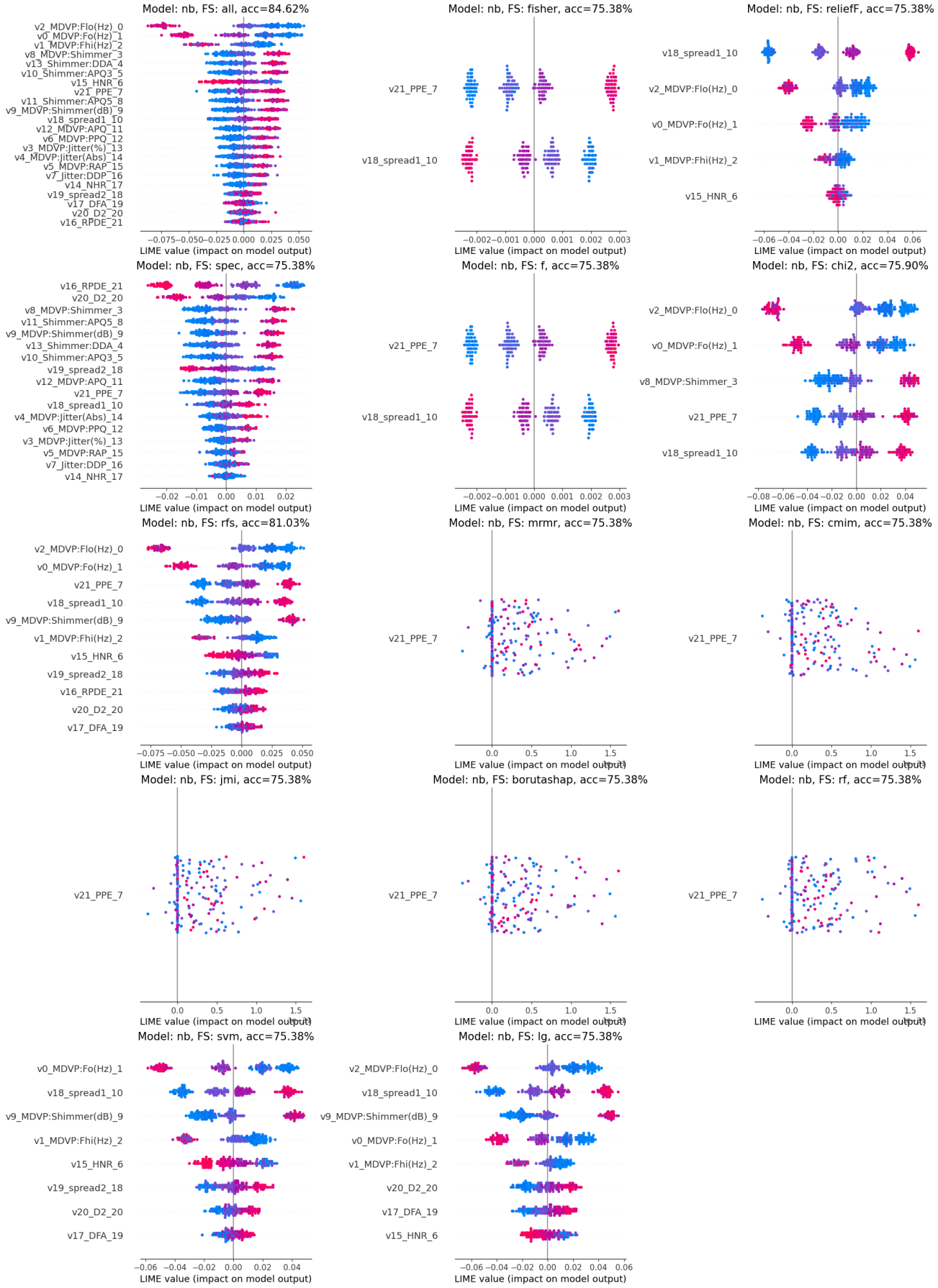


Fig. 11: Summary plots for the Oxford Parkinson's Disease Detection dataset in the *nb* model.

Fig. 12: Summary plots for the Oxford Parkinson's Disease Detection dataset in the *xg* model.

2) *Indian Liver Patient dataset:*

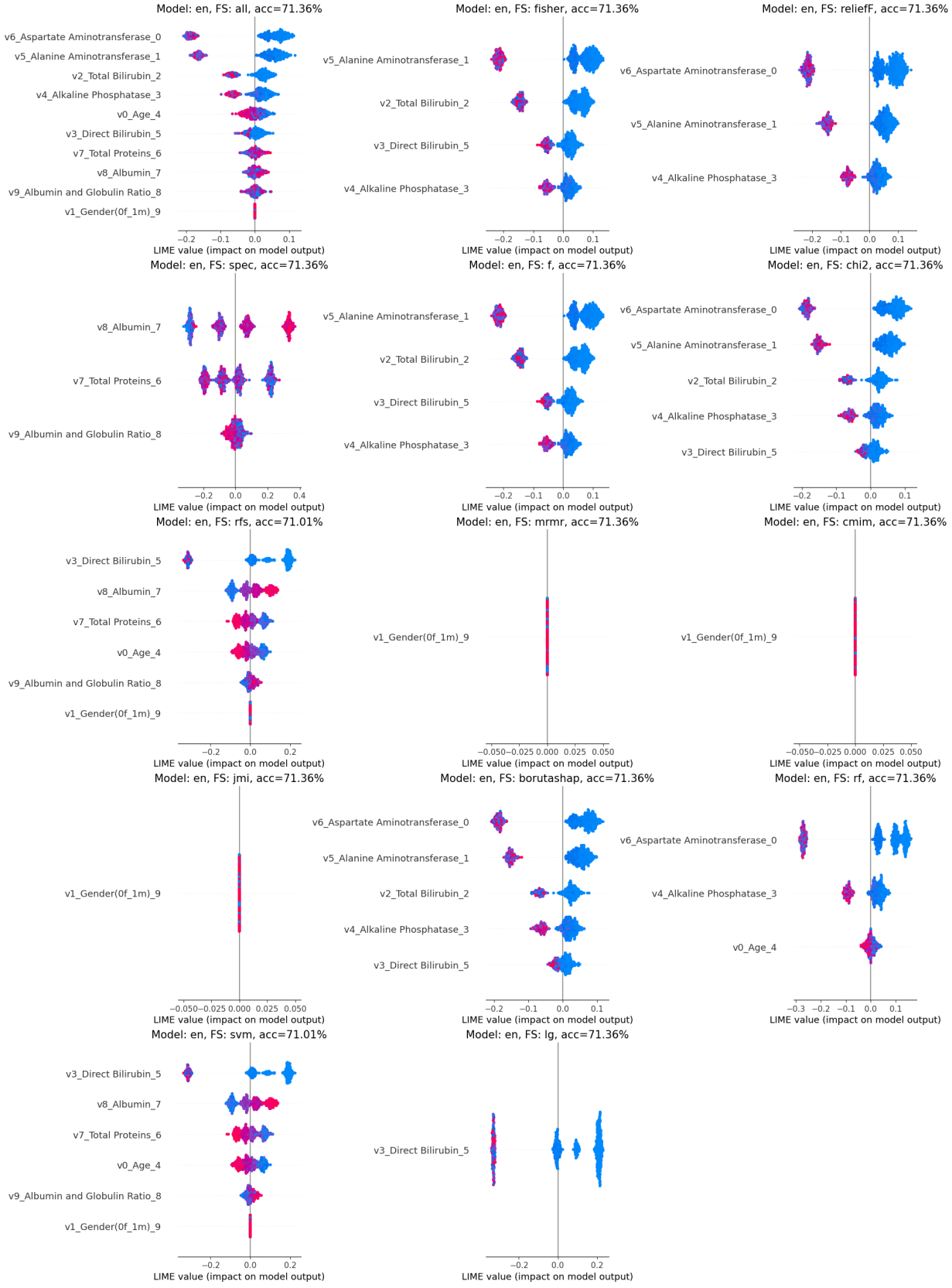


Fig. 13: Summary plots for the Indian Liver Patient dataset in the *en* model.

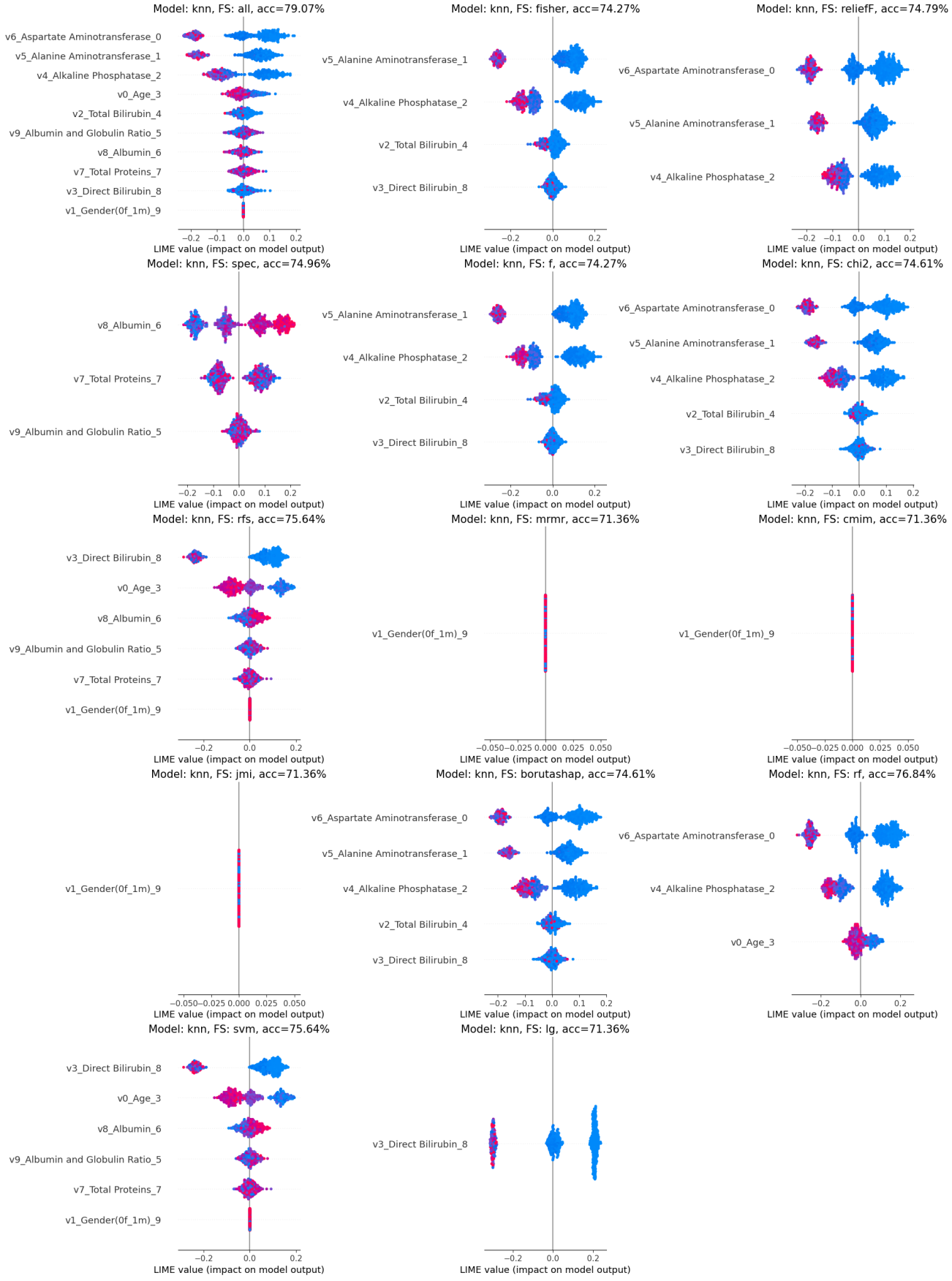


Fig. 14: Summary plots for the Indian Liver Patient dataset in the *knn* model.

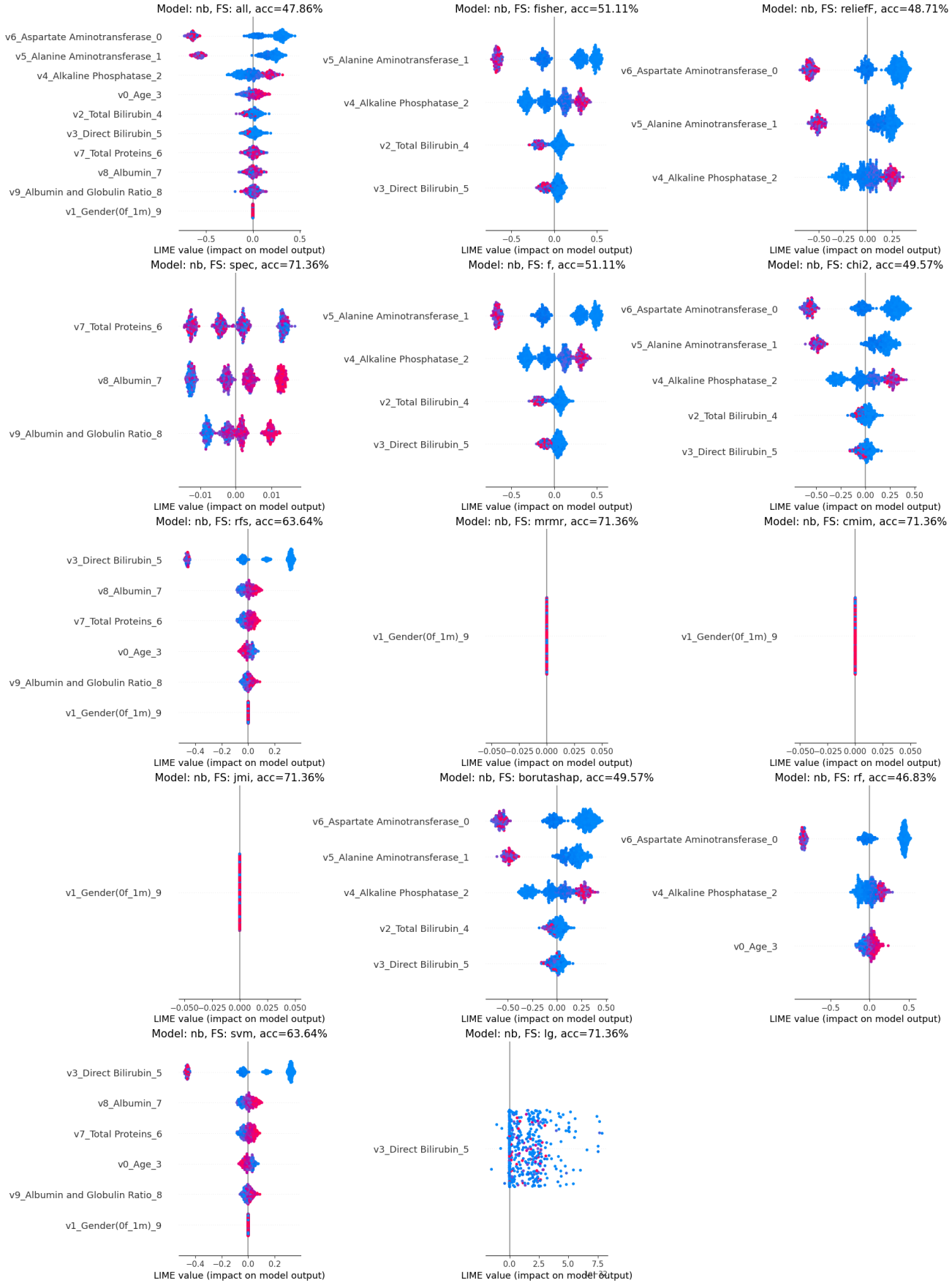


Fig. 15: Summary plots for the Indian Liver Patient dataset in the *nb* model.

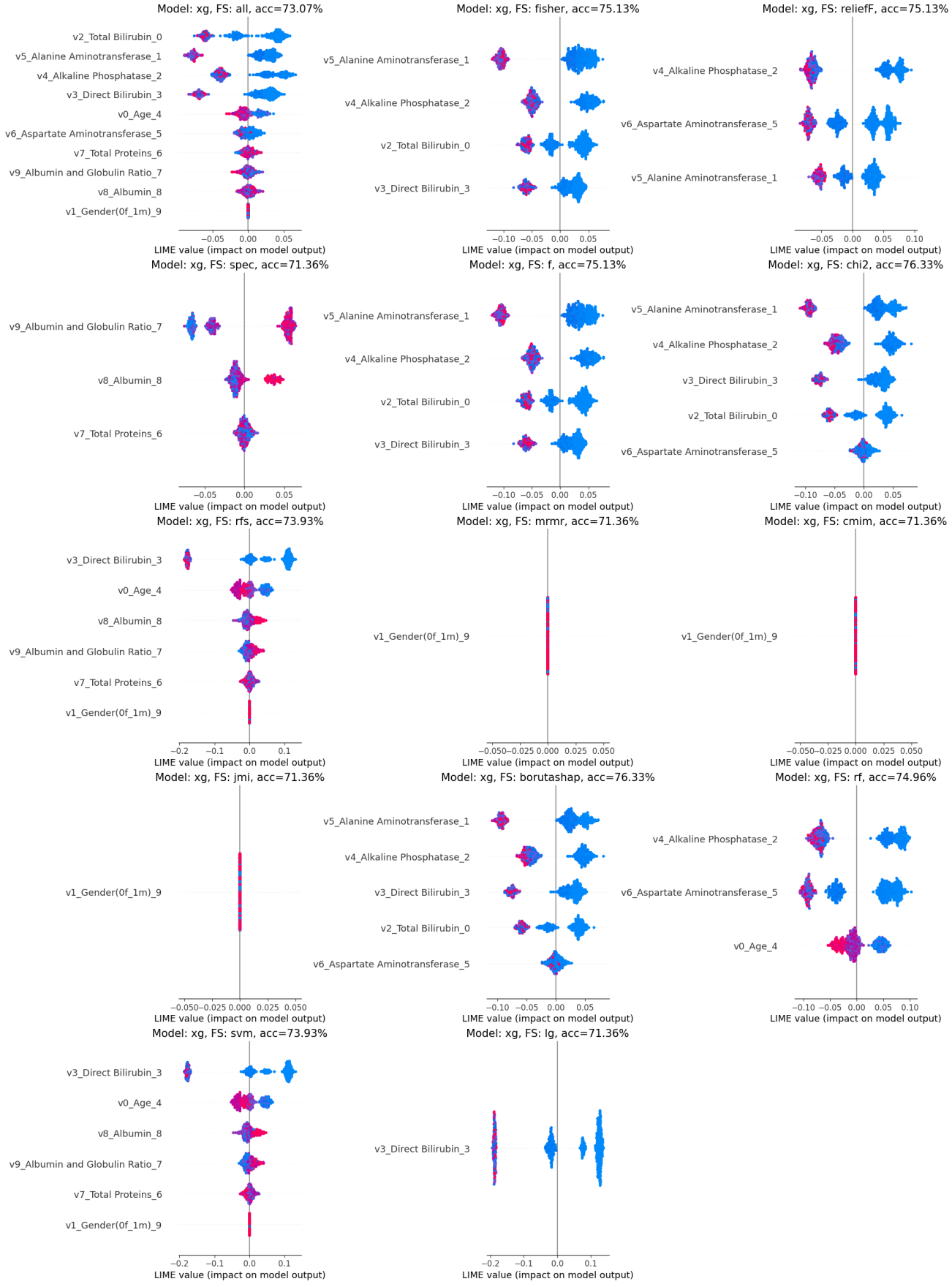


Fig. 16: Summary plots for the Indian Liver Patient dataset in the *xg* model.