Supplementary Material

Are machine learning based methods suited to address complex biological problems? Lessons from CAGI-5 challenges.

Castrense Savojardo[†], Giulia Babbi[†], Samuele Bovo[†], Emidio Capriotti[†], Pier Luigi Martelli[†,‡], Rita Casadio[†,§]

[†] Biocomputing Group, Department of Pharmacy and Biotechnology, University of Bologna, Italy.

[§] Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Italian National Research Council (CNR), Bari, Italy.

[‡] Corresponding author

Supplementary Tables

Table S1. INPS-3D predictions for the *FXN* SAV dataset

SAV	Experimental ΔΔG	Predicted ΔΔG (INPS-3D)
p.Asp104Gly	0.255	-0.776491
p.Ala107Val	0.22	-0.1439655
p.Ser202Phe	-0.685	0.4780699
p.Ser181Phe	-2.035	-0.115336
p.Phe109Leu	-2.645	-1.40127
p.Ser161Ile	-3.44	-0.697307
p.Tyr123Ser	-4.48	-2.577015
p.Trp173Cys	-9.5	-2.05509

Table S3. Predictions of different methods for the *CHEK2* SAV dataset

SAV	p_{case}	SNPs&GO	P_d	P_p	INPS
p.Ala584Val	1.00	0.760	0.560	0.750	0.840
p.Ala94Thr	1.00	0.590	0.540	0.594	0.500
p.Asp154Gly	1.00	0.880	0.820	0.654	0.793
p.Asp540Asn	1.00	0.530	0.650	0.570	0.500
p.Ser67Tyr	1.00	0.650	0.640	0.750	0.840
p.Glu547Gln	0.00	0.530	0.560	0.500	0.500
p.Glu64Lys	0.00	0.680	0.760	0.500	0.500
p.Leu555Val	0.50	0.560	0.540	0.834	0.500
p.Pro85Leu	0.00	0.500	0.590	0.714	0.775
p.Arg562Leu	0.00	0.650	0.960	0.500	0.500
p.Arg95Gln	0.00	0.650	0.750	0.500	0.500
p.Ser548Pro	0.00	0.680	0.690	0.500	0.500
p.Val66Met	0.00	0.650	0.630	0.630	0.500
p.Ile158Met	1.00	0.820	0.590	0.750	1.000
p.Ile200Thr	1.00	0.740	0.770	1.000	1.000
p.Ile203Arg	1.00	0.850	0.810	0.834	1.000
p.Arg160Gly	1.00	0.970	0.850	0.882	1.000
p.Arg180Gln	1.00	0.590	0.750	0.500	0.500
p.Arg191Gly	1.00	0.880	0.850	0.882	1.000
p.Arg223Cys	1.00	0.820	0.950	0.750	0.620
p.Asp481Tyr	1.00	0.760	0.910	0.630	0.500
p.Glu282Lys	1.00	0.760	0.760	0.500	0.500
p.Gly385Ser	1.00	0.850	0.820	0.500	0.673
p.Ile264Val	1.00	0.820	0.500	0.500	0.519
p.Leu279Pro	0.82	0.910	0.960	0.702	1.000
p.Asn448Lys	1.00	0.850	0.750	0.630	0.500

p.Pro527Leu	1.00	0.790	0.590	0.714	1.000
p.Arg449His	1.00	0.710	0.800	0.702	0.500
p.Thr519Met	1.00	0.760	0.600	0.500	0.804
p.Glu420Gly	0.00	0.760	0.660	0.870	1.000
p.Ile491Ser	0.33	0.680	0.900	1.000	1.000
p.Ile491Val	0.00	0.740	0.500	0.500	0.708
p.Arg389His	0.00	1.000	0.800	0.702	1.000
p.Thr421Ile	0.00	0.710	0.630	0.558	0.520

Table S4. Predictions of different methods for the *PCM1* SAV dataset

SAV	Functional effect	P_d (numeric)	P_d (class)	P_p (numeric)	P_p (class)	INPS (numeric)	INPS (class)
p.Gly6Asp	P	0.82	P	0.67	P	-0.767	H
p.Glu23Asp	B	0.35	B	0.43	H	-0.618	H
p.Thr77Ala	B	0.27	B	0.5	H	-0.299	B
p.Met146Val	H	0.46	H	0.57	H	-1.078	P
p.Ala156Val	H	0.45	H	0.6	P	0.494	B
p.Met200Thr	H	0.41	H	0.25	B	-1.666	P
p.Met200Ile	H	0.45	H	0.43	H	-0.138	B
p.Asp214Gly	P	0.66	P	0.52	H	-0.802	H
p.Glu248Gln	H	0.45	H	0.25	B	-0.556	H
p.Glu311Gln	H	0.45	H	0.25	B	-0.640	H
p.Glu369Gly	H	0.53	H	0.7	P	-1.134	P
p.Pro390Ser	H	0.48	H	0.5	H	-0.345	B
p.Leu472Val	H	0.43	H	0.67	P	-0.381	B
p.Gly482Val	P	0.82	P	0.7	P	-0.508	H
p.Glu543Lys	P	0.61	P	0.33	B	-0.300	B
p.Asp574Gly	P	0.66	P	0.52	H	-0.803	H
p.Arg604Leu	P	0.77	P	0.4	B	0.056	B
p.Glu624Lys	P	0.61	P	0.33	B	-0.225	B
p.Ile659Met	H	0.47	H	0.6	P	-1.423	P
p.Ser804Arg	P	0.62	P	0.43	H	-0.020	B
p.Arg833Thr	P	0.71	P	0.67	P	-0.976	H
p.Cys876Arg	P	0.76	P	0.72	P	-1.281	P
p.Gly892Trp	P	0.81	P	0.5	H	-0.780	H
p.Glu917Gly	H	0.53	H	0.7	P	-1.099	P
p.Lys954Asn	H	0.5	H	0.57	H	-0.755	H
p.Asn1125Ser	H	0.49	H	0.63	P	-0.296	B
p.Lys1275Glu	H	0.48	H	0.33	B	-0.594	H
p.His1352Tyr	H	0.58	H	0.42	H	-0.038	B
p.Cys1361Tyr	P	0.83	P	0.67	P	0.486	B
p.Ala1490Gly	H	0.47	H	0.58	H	-1.192	P
p.Glu1535Lys	P	0.61	P	0.33	B	-0.467	B
p.Ala1553Gly	H	0.47	H	0.58	H	-0.031	B
p.Gly1556Asp	P	0.82	P	0.67	P	-0.518	H
p.Lys1861Asn	H	0.5	H	0.57	H	-0.648	H
p.Asn1875Lys	P	0.6	P	0.5	H	-0.271	B
p.Arg1907His	P	0.64	P	0.56	H	0.052	B
p.Pro1913Leu	H	0.47	H	0.57	H	0.618	H
p.Ala1979Ser	B	0.33	B	0.27	B	-0.760	H