Supplementary Materials

Machine Learning Methods for Prediction of Food Effects on Bioavailability: A Comparison of Support Vector Machines and Artificial Neural Networks

Harriet Bennett-Lenane, Brendan T. Griffin, Joseph P. O'Shea.

School of Pharmacy, University College Cork, Cork, Ireland.

*Correspondence to: Joseph O' Shea; Tel.: +353 (0)21 4901665, Email address: joseph.oshea@ucc.ie

1. Results of Statistical Analysis comparing Positive, Negative and No FE groups for drugs licensed from 2016-2020
2. Visual comparison of the sensitivity, precision, specificity and Matthews correlation coefficient (MCC) performance metrics calculated for the test set of the optimum support vector machine (SVM) and artificial neural network (ANN) models produced in this study to predict FE classification.

1. Results of Statistical Analysis comparing Positive, Negative and No Food Effect (No FE) groups for drugs licensed from 2016-2020

Drug Property	Descriptors	Negative	No FE	Positive	Statistical Tests	Neg vs No FE	Neg vs Pos	No FE vs Pos
S+logP	*n*	17	80	44	Levene's Test	0.08	0.15	0.81
	Median	2.09	2.80	3.61	Bootstrap/t-test	0.018[B]	0.002[B]	0.021[t]
	Mean	1.50	2.90	3.67	Mean Difference	-1.41	-2.17	-0.76
	SD of Mean	2.29	1.72	1.77	95% Confidence Interval	(L) -2.56	(L) -3.41	(L) 1.41
	Q1, Q3,	-0.45, 3.33	2.02, 3.90	2.71, 4.80		(U) -0.34	(U) -0.98	(U) -0.11
	Min, Max	-2.43, 4.21	-1.38, 7.49	-1.34, 7.25				
	Variance	5.26	2.97	3.13				
Hydrogen	*n*	17	80	44	Levene's Test	0.075	0.26	0.34
Bond	*Median*	4	2	2	Bootstrap	0.00	0.00	0.86
Donors	*Mean*	3.53	1.90	1.86	Mean Difference	1.63	1.67	0.04
	SD of Mean	1.55	1.09	1.15	95% Confidence Interval	(L) 0.85	(L) 0.86	(L) -0.36
	Q1, Q3	2.5, 4.5	1, 3	1, 3		(U) 2.39	(U) 2.48	(U) 0.46
	Min, Max	0, 6	0, 4	0, 4				
	Variance	2.39	1.18	1.33				
Hydrogen	*n*	17	80	44	*Levene's Test*	*0.60*	*0.42*	*0.04*
Bond	*Median*	8	6	7	*Bootstrap*	*0.06*	*0.67*	*0.09*
Acceptors	*Mean*	7.71	6.28	7.30	*Mean Difference*	*1.43*	*0.41*	*-1.02*
	SD of Mean	2.69	2.57	3.49	*95% Confidence Interval*	*(L) -0.04*	*(L) -1.47*	*(L) -2.22*
	Q1, Q3	5.5, 10	5, 7.75	5, 9		*(U) 2.91*	*(U) 2.30*	*(U) 0.18*
	Min, Max	4, 13	1, 15	2, 16				
	Variance	7.221	6.61	12.17				
$logD_{7.4}$	*n*	17	80	44	Levene's Test	0.05	0.16	0.73
	Median	1.12	2.18	2.93	Bootstrap/t-test	0.09[B]	0.003[B]	0.003[t]
	Mean	0.82	1.98	3.01	Mean Difference	-1.16	-2.19	-1.03
	SD of Mean	2.65	1.78	1.93	95% Confidence Interval	(L) -2.65	(L) -3.66	(L) -1.71
	Q1, Q3	-0.57, 2.93	0.99, 3.13	1.75, 4,17		(U) 0.22	(U) -0.80	(U) -0.35
	Min, Max	-5.18, 4.21	-2.87, 6.64	-2.26, 7.25				
	Variance	7.07	3.17	3.71				
Polar Surface	*n*	17	80	44	Levene's Test	0.16	0.87	0.05
Area	*Median*	107.41	90.37	100.01	Bootstrap	0.00	0.07	0.047
	Mean	125.09	89.12	104.26	Mean Difference	35.92	20.77	-15.15
	SD of Mean	35.87	33.26	43.93	95% Confidence Interval	(L) 18.03	(L) -1.06	(L) -30.40
	Q1, Q3	92.88, 157.65	70.47, 108.48	77.36, 132.71		(U) 54.07	(U) 42.94	(U) 0.19
	Min, Max	76.10, 178.36	22.00, 188.80	29.46, 199.58				
	Variance	1286.99	1105.95	1930.20				
Dose (mg)	*n*	17	80	44	Levene's Test	0.25	0.009	0.258
	Median	80	50.00	200.00	Bootstrap	0.64	0.001	0.005
	Mean	105.36	121.15	242.57	Mean Difference	-15.79	-137.20	-121.41
	SD of Mean	96.00	205.71	193.88	95% Confidence Interval	(L) -79.20	(L) -212.56	(L) -195.66
	Q1, Q3	25.00, 175.00	10.00, 143.75	100.00, 375.00		(U) 48.38	(U) -62.23	(U) -44.74

Drug Property	Descriptors	Negative	No FE	Positive	Statistical Tests	Neg vs No FE	Neg vs Pos	No FE vs Pos
	Min, Max	1.00, 300.00	0.20, 1340.00	2.00, 800.00				
	Variance	9216.31	42314.80	37588.93				
S+Sw	*n*	17	80	44	Levene's Test	0.18	0.18	0.86
	Median	-0.49	-0.99	-1.54	Bootstrap/t-test	0.13B	0.003B	0.002t
	Mean	-0.38	-0.92	-1.56	Mean Difference	0.54	1.18	0.64
	SD	1.39	1.07	1.05	95% Confidence Interval	(L) -0.10	(L) 0.51	(L) 0.24
	Q1, Q3	-1.49, 0.22	-1.55, -0.31	-2.35, -0.76		(U) 1.25	(U) 1.91	(U) 1.03
	Min, Max	-2.05, 2.98	-3.61, 2.02	-4.42, 1.01				
	Variance	1.94	1.14	1.097				
MAD	*n*	17	80	44	Levene's Test	0.07	0.07	0.78
	Median	2.46	2.78	1.02	t-test	0.96	0.06	0.001
	Mean	2.77	2.75	2.11	Mean Difference	0.15	0.67	0.64
	SD	1.54	1.06	1.02	95% Confidence Interval	(L) -0.60	(L) -0.02	(L) 0.25
	Q1, Q3	1.57, 3.54	2.15, 3.38	1.26, 2.71		(U) 0.63	(U) 1.33	(U) 1.03
	Min, Max	0.57, 6.14	0.02, 5.84	-0.13, 4,34				
	Variance	2.36	1.13	1.04				
Molecular Weight (g/mol)	*n*	17	80	44	Levene's Test	0.09	0.86	0.04
	Median	457.69	426.31	446.95	Bootstrap	0.47	0.41	0.02
	Mean	453.74	419.89	497.65	Mean Difference	33.85	-43.91	-77.78
	SD of Mean	184.44	134.86	180.56	95% Confidence Interval	(L) -59.07	(L) -147.60	(L) -141.53
	Q1, Q3	302.62, 570.61	348.99, 491.83	368.95, 557.87		(U) 126.99	(U) 60.81	(U) -22.05
	Min, Max	146.24, 804.04	109.13, 889.02	287.22, 1113.21				
	Variance	34018.56	18188,43	32601.00				
Dose/Solubility Ratio	*n*	17	80	44	Levene's Test	0.57	0.26	0.28
	Median	2.14	2.47	3.86	t-test	0.28	0.00	0.00
	Mean	2.10	2.47	3.71	Mean Difference	-0.37	-1.61	-1.24
	SD of Mean	1.36	1.23	1.20	95% Confidence Interval	(L) -1.03	(L) -2.32	(L) -1.70
	Q1, Q3	1.08, 3.22	1.62, 3.26	3.32, 4.40		(U) 0.30	(U) -0.90	(U) -0.79
	Min, Max	-0.89, 4.05	-0.54, 5.83	0.18, 7.02				
	Variance	1.86	1.52	1.43				
Rotatable Bonds	*n*	17	80	44	Levene's Test	0.15	0.89	0.02
	Median	6	5	5	Bootstrap	0.26	0.96	0.20
	Mean	6.65	5.50	6.60	Mean Difference	1.15	0.079	-1.07
	SD of Mean	5.26	3.41	4.97	95% Confidence Interval	(L) -0.86	(L) -2.81	(L) -2.79
	Q1, Q3	2.50, 8.50	3.23, 7.75	3.00, 9.00		(U) 3.15	(U) 2.97	(U) 0.49
	Min, Max	0.00, 22.00	0.00, 15.00	0.00, 24.00				
	Variance	27.62	11.60	24.72				

2. Visual comparison of the sensitivity, precision, specificity and Matthews Correlation Coefficient (MCC) performance metrics calculated for the test set of the optimum Support Vector Machines (SVM) and Artificial Neural Network (ANN) models produced in this study to predict food effect (FE) classification.