

A Multi-Task Learning Approach for Prediction of Missing Bioactivity Values of Compounds for the SLC Transporter Superfamily

Tarik Čerimagić, Sergey Sosnin, and Gerhard F. Ecker*

University of Vienna, Department of Pharmaceutical Sciences

Josef Hlaubek Platz 2, 1090 Vienna, Austria

Address correspondence to: gerhard.f.ecker@univie.ac.at

Supplementary information

Supplementary Table 1 Hyperparameter ranges and categories set for the Bayesian search of the single-task deep neural network (ST-DNN) models. Left column shows the used parameters and right column shows their corresponding ranges [] or categories ().

Hyperparameter	Suggested value range-[] / category-()
nr_of_shared_layers	(2, 3)
hidden_size_1	[100-1200]
hidden_size_2	[5-500]
hidden_size_3	[2-500] if nr_of_shared_layers = 3
learning_rate	[0.000035-0.000170]
dropout rate	[0.07-0.50]
batch_size	[8, 16, 32, 64]
patience	[7-10]

Supplementary Table 2 Hyperparameter ranges and categories set for the Bayesian search of the single-task random forest (ST-RF) models. Left column shows the used parameters and right column shows their corresponding ranges [] or categories ().

Hyperparameter	Suggested value range-[] / category-()
n_estimators	[50-300]
max_depth	[5-50]
min_samples_split	[2-20]
min_samples_leaf	[1-10]
max_features	(None, "sqrt", "log2")
bootstrap	(True, False)
max_samples	[0.5-1.0] if bootstrap = True, else None

Supplementary Table 3 UMAP parameters for 2D chemical space of SLC compounds

Hyperparameter	Value
metric	"euclidean"
neighbors	200
n_components	2
low_memory	False
min_dist	0.99

Supplementary Table 4 Hyperparameters for data imputation models. First column contains the names of imputation models, second column contains specific hyperparameters that were optimized, and third column contains categories of each parameter explored during the grid search.

Imputation model	Hyperparameter	Suggested category-[]
SimpleImputer	strategy	['mean', 'median', 'most_frequent', 'constant']
KNNImputer	neighbors	[2, 4, 5, 10],
	'weights'	['uniform', 'distance']
IterativeImputer	max_iter	[5, 10, 20],
	'sample_posterior'	[False, True]
SoftImpute	shrinkage_value	[None, 0.1, 0.5, 1.0]
	convergence_threshold	[1e-3, 1e-4, 1e-5]
	max_iters	[100, 200],
	max_rank"	[None, 50, 100],,]
IterativeSVD	rank	[10, 50, 100]
	convergence_threshold	[1e-4, 1e-5]
	max_iters	[200, 400]
MatrixFactorization	rank	[10, 40, 100]
	learning_rate	[0.001, 0.01, 0.05]
	max_iters	[500, 1000],
	convergence_threshold	[1e-4, 1e-5]

Supplementary Table 5 Target-based results. There are 31 unique targets in the first column. Each target has number of compounds shown in the second column, number of total members belonging to the same family in the third column, and a number compounds a shared family of targets has in total in fourth column. The performance scores for multi-task deep neural network (MT-DNN), ST-DNN, and ST-RF are shown in the fifth, sixth, and seventh column respectively. Each performance score (Q^2) acquired from 5-fold cross-validation is followed by the standard deviation (SD) value.

Target	Compounds	Family members	Total compounds	MT-DNN (Q^2)	ST-DNN (Q^2)	ST-RF (Q^2)
SLC1A5	23	4	297	-1.888 ± 3.956	-6.160 ± 7.902	-0.707 ± 0.554
SLC1A1	74	4	297	0.444 ± 0.170	0.116 ± 0.284	0.198 ± 0.037
SLC1A2	91	4	297	0.532 ± 0.370	0.487 ± 0.349	0.555 ± 0.117
SLC1A3	109	4	297	0.239 ± 0.323	0.211 ± 0.162	-0.261 ± 0.475
SLC2A1	562	3	685	0.385 ± 0.126	0.293 ± 0.178	0.478 ± 0.062
SLC2A2	34	3	685	-0.395 ± 0.461	-1.141 ± 1.054	-0.210 ± 0.568
SLC2A3	89	3	685	0.017 ± 0.159	-0.682 ± 0.652	-0.069 ± 0.264
SLC5A7	18	4	2194	-0.487 ± 0.577	-9.773 ± 4.424	-0.277 ± 0.330
SLC5A2	1173	4	2194	0.698 ± 0.030	0.696 ± 0.052	0.678 ± 0.046
SLC5A1	975	4	2194	0.809 ± 0.006	0.795 ± 0.021	0.796 ± 0.023
SLC5A4	28	4	2194	0.758 ± 0.107	0.428 ± 0.587	0.610 ± 0.210
SLC6A4	2254	10	6013	0.587 ± 0.042	0.607 ± 0.049	0.545 ± 0.037
SLC6A12	15	10	6013	-271.183 ± 603.939	-37.855 ± 70.495	-2.276 ± 5.065
SLC6A11	23	10	6013	-0.807 ± 1.112	-4.423 ± 5.174	-0.805 ± 1.010
SLC6A15	28	10	6013	-0.116 ± 0.695	-0.810 ± 0.442	-0.460 ± 0.806

SLC6A1	43	10	6013	0.009 ± 0.349	-1.581 ± 1.341	-0.159 ± 0.321
SLC6A7	49	10	6013	0.150 ± 0.220	0.081 ± 0.500	0.012 ± 0.334
SLC6A2	1616	10	6013	0.620 ± 0.027	0.582 ± 0.043	0.553 ± 0.045
SLC6A3	1304	10	6013	0.617 ± 0.016	0.621 ± 0.064	0.535 ± 0.038
SLC6A9	574	10	6013	0.562 ± 0.095	0.523 ± 0.094	0.542 ± 0.064
SLC6A5	107	10	6013	0.477 ± 0.250	0.193 ± 0.297	0.356 ± 0.112
SLC7A11	22	2	48	0.763 ± 0.171	-1.587 ± 2.862	0.309 ± 0.371
SLC7A5	26	2	48	-2.496 ± 5.524	-9.555 ± 10.105	-0.071 ± 0.485
SLC8A1	320	1	320	0.257 ± 0.108	0.001 ± 0.275	0.179 ± 0.074
SLC9A1	236	2	501	0.558 ± 0.231	0.570 ± 0.072	0.530 ± 0.041
SLC9A3	265	2	501	0.471 ± 0.093	0.270 ± 0.182	0.086 ± 0.251
SLC10A2	231	2	278	0.542 ± 0.059	0.452 ± 0.238	0.453 ± 0.153
SLC10A1	47	2	278	0.218 ± 0.680	0.062 ± 1.015	0.446 ± 0.233
SLC11A2	42	1	42	-1.184 ± 2.043	-5.065 ± 7.916	-0.110 ± 0.266
SLC15A1	14	1	14	-53.868 ± 49.803	-6005.621 ± 7229.414	-962.730 ± 1100.560
SLC16A3	12	1	12	-1.992 ± 3.989	-34.527 ± 37.945	-22.157 ± 32.519
SLC18A3	79	2	98	-0.041 ± 0.497	0.008 ± 0.538	-0.126 ± 0.319
SLC18A2	19	2	98	-7.474 ± 12.709	-8.726 ± 6.474	-0.660 ± 1.339
SLC19A1	22	1	22	-0.949 ± 1.683	-66.792 ± 108.516	-10.617 ± 15.402
SLC22A1	84	6	669	-0.383 ± 0.377	-1.305 ± 1.574	-0.210 ± 0.174
SLC22A2	54	6	669	-0.631 ± 0.506	-2.586 ± 4.034	-2.195 ± 2.326
SLC22A8	22	6	669	-2.264 ± 3.569	-12.050 ± 8.857	-1.280 ± 1.187
SLC22A6	40	6	669	-0.367 ± 0.320	-0.631 ± 1.256	0.047 ± 0.158
SLC22A3	26	6	669	-1.605 ± 1.283	-1.836 ± 2.333	-0.750 ± 0.667
SLC22A12	443	6	669	0.554 ± 0.070	0.521 ± 0.084	0.560 ± 0.026
SLC27A1	38	2	65	-0.259 ± 0.688	-2.308 ± 1.320	-0.564 ± 0.939
SLC27A4	27	2	65	-0.138 ± 1.093	-7.240 ± 14.314	-17.641 ± 33.550
SLC28A2	35	1	35	0.159 ± 0.382	-1.325 ± 0.777	-1.504 ± 1.449
SLC29A1	140	1	140	0.523 ± 0.153	0.365 ± 0.204	0.496 ± 0.107
SLC33A1	131	1	131	-2.554 ± 1.548	-2.866 ± 3.151	0.110 ± 0.133
SLC34A2	68	1	68	0.050 ± 0.671	-0.595 ± 1.552	-0.052 ± 0.294
SLC37A4	17	1	17	-0.155 ± 1.021	-5.303 ± 6.055	-0.084 ± 1.141
SLC40A1	397	1	397	0.398 ± 0.106	0.407 ± 0.126	0.483 ± 0.050
SLC46A1	28	1	28	-1.281 ± 1.689	-2.922 ± 3.610	-1.938 ± 2.775
SLC47A1	42	2	64	-0.637 ± 0.461	-5.216 ± 2.823	-0.553 ± 0.777
SLC47A2	22	2	64	-6.472 ± 8.538	-8.541 ± 6.298	-0.218 ± 0.387
SLC65A2	97	1	97	0.188 ± 0.535	-0.138 ± 0.495	0.292 ± 0.182
SLCO1B1	128	2	220	0.062 ± 0.364	-0.432 ± 0.791	-0.541 ± 0.271
SLCO1B3	92	2	220	-0.666 ± 0.564	-0.820 ± 0.828	-0.013 ± 0.151