Model Construction and Performance

Data was extracted from ChEMBL (https://www.ebi.ac.uk/chembl/, version 23) and ToxCast (https://www.epa.gov/chemical-research/toxicity-forecasting).

Chemical fingerprints were calculated, and models constructed using the fingerprint generator and model generator included in this location. Models developed in TensorFlow during the research (1) and improved models optimized in keras (2) are included. The best performing models included are those optimized in keras. The models included and newly generated models can be recalled using the model recaller.

Model performance is shown below.

Models (1)

Train	TP	FN	TN	FP	SE	SP	ACC	MCC	ROC-AUC
SERT [10,10] RR = 0.01, LR = 1.0, i = 1000	3060	0	0	823	100.0	0.0	78.8	N/A	0.50
hERG [10,10] RR = 0.01, LR = 0.1, i = 3000	3662	0	0	2443	100.0	0.0	60.0	N/A	0.50
ADORA2A [10,10] RR = 0.01, LR = 0.5, i = 1000	2795	164	1367	193	94.5	87.6	92.1	0.825	0.95
AR [1000] RR = 0.1, LR = 0.025, i = 2000	1187	782	5343	128	60.3	97.7	87.8	0.670	0.85
AChE [1000] RR = 0.75, LR = 0.01, i = 1800	1743	220	1212	257	88.8	82.5	86.1	0.715	0.91
AVERAGE					88.7	53.6	81.0	N/A	0.74

Test	TP	FN	TN	FP	SE	SP	ACC	MCC	ROC-AUC
SERT [10,10] RR = 0.01, LR = 1.0, i = 1000	983	0	0	312	100.0	0.0	75.9	N/A	0.50
hERG [10,10] RR = 0.01, LR = 0.1, i = 3000	1233	0	0	803	100.0	0.0	60.6	N/A	0.50
ADORA2A [10,10] RR = 0.01, LR = 0.5, i = 1000	941	43	425	98	95.6	81.3	90.6	0.791	0.93
AR [1000] RR = 0.1, LR = 0.025, i = 2000	409	259	1769	44	61.2	97.6	87.8	0.675	0.84
AChE [1000] RR = 0.75, LR = 0.01, i = 1800	560	91	382	112	86.0	77.3	82.3	0.637	0.88
AVERAGE					88.6	51.2	79.4	N/A	0.73

Models (2)

Train	TP	FN	TN	FP	SE	SP	ACC	MCC	ROC-AUC
SERT [10,10] RR = 0.001, LR = 0.01, i = 100	3026	10	842	5	99.7	99.4	99.6	0.989	1.00
hERG [10,10] RR = 0.05, LR = 0.001, i = 200	3420	221	1897	567	93.9	77.0	87.1	0.731	0.94
ADORA2A [10,10] RR = 0.001, LR = 0.0001, i = 200	2940	18	1539	22	99.4	98.6	99.1	0.980	1.00
AR [1000] RR = 0.001, LR = 0.00001, i = 500	1408	544	5442	46	72.1	99.2	92.1	0.791	0.94
AChE [1000] RR = 0.01, LR = 0.0001, i = 100	1658	299	425	63	84.7	87.1	85.2	0.629	0.94

AVERAGE							90.0	92.2	92.6	0.8		0.94
Test		**TP**	**FN**	**TN**	**FP**		**SE**	**SP**	**ACC**	**MCC**		**ROC-AUC**
SERT [10,10] RR = 0.001, LR = 0.01, i = 100		980	27	264	24		97.3	91.7	96.1	0.887		0.98
hERG [10,10] RR = 0.05, LR = 0.001, i = 200		1090	164	499	283		86.9	63.8	78.0	0.527		0.85
ADORA2A [10,10] RR = 0.001, LR = 0.0001, i = 200		957	28	486	36		97.2	93.1	95.8	0.906		0.99
AR [1000] RR = 0.001, LR = 0.00001, i = 500		453	232	1771	25		66.1	98.6	89.6	0.734		0.87
AChE [1000] RR = 0.01, LR = 0.0001, i = 100		539	118	425	63		82.0	87.1	84.2	0.685		0.91
AVERAGE							85.9	86.9	88.7	0.7		0.92

Models and source data are included for each biological target, the serotonin transporter (SERT), human ether-à-go-go related gene channel (hERG), adenosine A2a receptor (ADORA2A), androgen receptor (AR) and acetylcholinesterase (AChE). Model architectures are reported in square brackets with commas separating hidden layers. Hyperparameters are reported as regularization rate (RR), learning rate (LR), and number of training iterations (i).

Models were evaluated based on the number of true positives (TP), false negatives (FN), true negatives (TN) and false positives (FP) on both the training and test sets. These values were used to calculate evaluation statistics; sensitivity (SE), specificity (SP), accuracy (ACC) and Matthews correlation coefficient (MCC).

$$SE = \frac{TP}{TP + FN}$$

$$SP = \frac{TN}{TN + FP}$$

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

A receiver operating curve (ROC) was plotted and the area under this curve (ROC-AUC) calculated.

Network activation similarity values are calculated between the input molecules and the training set as part of the recall procedure.