“A comparison of different compound representations for drug sensitivity prediction” - Supplementary Material

Table S1 Hyperparameter values considered for ECFP4, ECFP6, RDKitFP, AtomPair and LayeredFP models.

Hyperparameter	Values considered
Hidden layer sizes	[512, 256], [256, 128], [128, 64], [64, 32], [512, 256, 128], [256, 128, 64], [128, 64, 32], [512, 256, 128, 64], [256, 128, 64, 32]
L2 regularization	0, 0.0001, 0.001, 0.01
Batch normalization	True, False
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S2 Hyperparameter values considered for MACCS keys models.

Hyperparameter	Values considered
Hidden layer sizes	[128, 64], [64, 32], [32, 16], [16, 8], [128, 64, 32], [64, 32, 16], [32, 16, 8], [128, 64, 32, 16], [64, 32, 16, 8]
L2 regularization	0, 0.0001, 0.001, 0.01
Batch normalization	True, False
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S3 Hyperparameter values considered for Mol2vec models.

Hyperparameter	Values considered
Hidden layer sizes	[256, 128], [128, 64], [64, 32], [32, 16], [256, 128, 64], [128, 64, 32], [64, 32, 16], [256, 128, 64, 32], [128, 64, 32, 16]
L2 regularization	0, 0.0001, 0.001, 0.01
Batch normalization	True, False
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S4 Hyperparameter values considered for TextCNN models.

Hyperparameter	**Values considered**
Length of embedding vector	75, 32, 64
Kernel sizes	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20], [1, 2, 3, 4, 5, 7, 10, 15], [3, 4, 5, 7, 10, 15], [3, 4, 5, 7, 10], [3, 4, 5, 7], [3, 4, 5], [3, 5, 7]
Number of filters	[100, 200, 200], [32, 32, 32], [128, 128, 128], [100, 200, 200, 200], [32, 32, 32, 32], [128, 128, 128, 128], [100, 200, 200, 200, 200], [32, 32, 32, 32, 64], [128, 128, 128, 128, 64], [100, 200, 200, 200, 200, 100], [32, 32, 32, 32, 64, 64], [128, 128, 128, 128, 64, 64], [100, 200, 200, 200, 200, 100, 100, 100], [32, 32, 32, 32, 64, 64, 64, 64], [128, 128, 128, 128, 64, 64, 64, 64], [100, 200, 200, 200, 200, 100, 100, 100, 100, 100, 160, 160], [32, 32, 32, 32, 64, 64, 64, 64, 128, 128, 128, 128], [128, 128, 128, 128, 64, 64, 64, 64, 32, 32, 32, 32]
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S5 Hyperparameter values considered for GraphConv models.

Hyperparameter	**Values considered**
Graph convolution layers	[32, 32], [64, 64], [128, 128], [32, 32, 32], [64, 64, 64], [128, 128, 128], [32, 32, 32, 32], [64, 64, 64, 64], [128, 128, 128, 128]
Dense layer units	2048, 1024, 512, 256, 128, 64, 32
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S6 Hyperparameter values considered for GCN models.

Hyperparameter	**Values considered**
Graph convolution layers	[32, 32], [64, 64], [128, 128], [32, 32, 32], [64, 64, 64], [128, 128, 128], [32, 32, 32, 32], [64, 64, 64, 64], [128, 128, 128, 128]
Predictor hidden units	256, 128, 64
Dropout rate	0, 0.25, 0.5
Predictor dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S7 Hyperparameter values considered for GAT models.

Hyperparameter	**Values considered**
Graph attention layers	[8, 8], [16, 16], [32, 32], [8, 8, 8], [16, 16, 16], [32, 32, 32]
Number of attention heads	4, 8
Predictor hidden units	256, 128, 64
Dropout rate	0, 0.25, 0.5
Predictor dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S8 Hyperparameter values considered for AttentiveFP models.

Hyperparameter	**Values considered**
Number of layers	1, 2, 3, 4
Number of time steps	2, 3, 4
Graph features size	32, 64, 128, 256, 512, 200
Dropout rate	0, 0.25, 0.5
Learning rate	0.0001, 0.001, 0.01

Table S9 Best hyperparameters for ECFP4 models for each of the datasets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[512, 256, 128, 64]	[256, 128, 64, 32]	[128, 64, 32]	[128, 64, 32]	[256, 128, 64, 32]
L2 regularization	0	0.0001	0.001	0.001	0.0001
Batch normalization	True	True	True	True	True
Dropout rate	0	0.5	0.25	0.25	0.5
Learning rate	0.001	0.01	0.01	0.01	0.01

Table S10 Best hyperparameters found for ECFP6 models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[256, 128, 64, 32]	[512, 256, 128]	[128, 64, 32]	[128, 64, 32]	[256, 128, 64, 32]
L2 regularization	0.0001	0.0001	0.001	0.001	0.0001
Batch normalization	True	True	True	True	True
Dropout rate	0.5	0.5	0.25	0.25	0.5
Learning rate	0.01	0.0001	0.01	0.01	0.01

Table S11 Best hyperparameters for MACCS keys models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[64, 32, 16, 8]	[64, 32, 16, 8]	[128, 64]	[128, 64]	[64, 32]
L2 regularization	0	0	0.001	0.001	0.0001
Batch normalization	True	True	False	False	False
Dropout rate	0.25	0.25	0	0	0.25
Learning rate	0.001	0.001	0.001	0.001	0.001

Table S12 Best hyperparameters for RDKitFP models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[512, 256]	[512, 256, 128]	[512, 256]	[512, 256]	[256, 128]
L2 regularization	0.0001	0	0.001	0.001	0.01
Batch normalization	False	True	False	False	False
Dropout rate	0.5	0.25	0	0	0
Learning rate	0.0001	0.0001	0.001	0.001	0.0001

Table S13 Best hyperparameters for AtomPair models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[512, 256]	[512, 256]	[512, 256]	[512, 256]	[256, 128]
L2 regularization	0.0001	0.0001	0.001	0.001	0.0001
Batch normalization	False	False	False	False	False
Dropout rate	0.5	0.5	0	0	0.25
Learning rate	0.0001	0.0001	0.001	0.001	0.001

Table S14 Best hyperparameters for LayeredFP models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[512, 256]	[512, 256]	[256, 128]	[256, 128]	[256, 128]
L2 regularization	0.0001	0.0001	0.01	0.01	0.01
Batch normalization	False	False	False	False	False
Dropout rate	0.5	0.5	0	0	0
Learning rate	0.0001	0.0001	0.0001	0.0001	0.0001

Table S15 Best hyperparameters for Mol2vec models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Hidden layer sizes	[256, 128, 64, 32]	[32, 16]	[64, 32, 16]	[256, 128, 64]	[256, 128, 64, 32]
L2 regularization	0	0.001	0.01	0.01	0
Batch normalization	True	False	True	True	True
Dropout rate	0	0	0.5	0.5	0
Learning rate	0.001	0.001	0.01	0.01	0.001

Table S16 Best hyperparameters for TextCNN models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Length of embedding vector	75	32	75	64	75
Kernel sizes	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]	[3, 4, 5, 7, 10, 15]	[1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20]	[3, 4, 5, 7, 10, 15]
Number of filters	[100, 200, 200, 200, 200, 100, 100, 100, 100, 100, 160, 160]	[100, 200, 200, 200, 200, 100, 100, 100, 100, 100, 160, 160]	[100, 200, 200, 200, 200, 100]	[100, 200, 200, 200, 200, 100, 100, 100, 100, 100, 160, 160]	[100, 200, 200, 200, 200, 100]
Dropout rate	0	0.25	0.5	0.25	0.5
Learning rate	0.001	0.0001	0.001	0.001	0.001

Table S17 Best hyperparameters for GraphConv models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Graph convolution layers	[32, 32, 32]	[64, 64]	[64, 64, 64]	[64, 64, 64]	[64, 64, 64]
Dense layer units	2048	256	512	512	512
Dropout rate	0	0	0	0	0
Learning rate	0.0001	0.01	0.001	0.001	0.001

Table S18 Best hyperparameters for GCN models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Graph convolution layers	[128, 128]	[128, 128, 128, 128]	[128, 128, 128, 128]	[32, 32]	[128, 128, 128, 128]
Predictor hidden units	128	64	64	64	64
Dropout rate	0	0	0	0	0
Predictor dropout rate	0	0.25	0.25	0	0.25
Learning rate	0.0001	0.001	0.001	0.001	0.001

Table S19 Best hyperparameters for GAT models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Graph attention layers	[32, 32]	[16, 16, 16]	[16, 16]	[32, 32]	[16, 16]
Number of attention heads	4	8	8	4	8
Predictor hidden units	256	256	64	64	128
Dropout rate	0	0	0.25	0	0
Predictor dropout rate	0.25	0.25	0	0.5	0.25
Learning rate	0.01	0.001	0.01	0.001	0.001

Table S20 Best hyperparameters for AttentiveFP models for each of the data sets used in this study.

Hyperparameter	NCI 1	NCI 109	PC-3	CCRF-CEM	A549/ATCC
Number of layers	1	1	2	1	1
Number of time steps	4	4	4	3	4
Graph features size	256	256	128	200	256
Dropout rate	0.25	0.25	0.25	0	0.25
Learning rate	0.001	0.001	0.01	0.01	0.001