A comprehensive evaluation of regression-based drug responsiveness prediction models, using cell viability inhibitory concentrations (IC50 values)

Park *et al*.

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Information

Supplementary Method S1.

Supplementary Tables S1 through S23

Supplementary Figures S1 through S5

Supplementary References

Supplementary Method S1.

Raw data description

For learning data, mutation profiles of cancer cell lines were obtained from the GDSC, consisting of 21,213 mutation sites, in 1001 cell lines, and the CCLE, providing gene expression mRNA profiles, of 18,988 genes, in 1,037 cell lines. DNA methylation was also obtained from CCLE, including a total of 20,192 positions in promotor regions located 1kb upstream of the transcription start sites of 16,493 genes, in 843 cancer cell lines. Copy number variation from CCLE was obtained, and it includes 17,006 genomic sites, in 1,043 cell lines. PaDEL software (Yap, 2011) generated chemical properties of drugs, from the databases, by taking drug SMILES (simplified molecular-input line-entry system) format as input.

Scenarios of learning dataset construction

For 21,213 mutation sites, mutational events were binarized to either 1 (presence) or 0 (absence), for cancer cell lines. Gene expression profiles ("E") of the cancer cell lines in the CCLE were standardized through z-normalization.

For DNA methylation ("Y"), the missing values, for each methylation site, was imputed by the 'impute' package from Bioconductor (Hastie T, 2019), and, subsequently, DNA methylation values were standardized through Z-normalization. Copy number variation ("O") data were used as its raw values.

The individual features, in drug chemical property data, were processed using the min-max normalization method. Drug response ln(IC50) values from the GDSC were used as output values for model learning. The schematic structure of input data is shown in Fig. 2a and Supplementary Figure S1.

ML model construction, according to dataset scenarios

It is noted that, in scenario 1, lasso and ridge were not used, due to extremely slow computation time. Each ML method used the scikit-learn python package. In scenario 1, optimal hyperparameters for SVR, random forest, and XGBoost were selected while searching for the best performance in a set of hyperparameter values (Supplementary Tables S19, S20, and S21). As a result, we selected

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

0.1 as the regularization parameter C in SVR; 200 as the number of trees in random forest; and 150 and 8 as the number of trees and the depth of trees, respectively, in XGBoost (Supplementary Table S18). In scenarios 2 through 11, ridge and lasso were trained under alpha 0.001; SVR under C 0.01; and random forest and XGBoost under default options (Supplementary Table S18).

Application of the ResNetIC50 model to clinical follow-up observations of chemotherapeutic agents in The Cancer Genome Atlas (TCGA) breast cancer (henceforth, TCGA-BRCA) dataset

From the TCGA-BRCA patients (Cancer Genome Atlas Research Network, 2012), we selected 76 patients under two criteria: (i) patients with at least one mutation; and (ii) patients treated with one out of four chemotherapeutic agents (docetaxel, doxorubicin, gemcitabine, and paclitaxel). In the drug response record of the clinical follow-up data, complete response and stable disease were regarded as responders and clinical progressive disease as non-responders.

In the follow-up, 28 of the 76 patients were treated with docetaxel, with 28 responders and one non-responder. 29 were given doxorubicin, with all responders. One patient was treated by gemcitabine, as a non-responder. 18 were given paclitaxel, all responders.

For predicting ln(IC50) values, for individual patients, his/her mutation status and drug chemical property vectors of his/her chemotherapeutic administration were concatenated, for input into ResNetIC50. ResNetIC50-predicted ln(IC50) values were classified under the criterion: IC50 < 1 μM (ln(IC50) < 0) for responders, and otherwise for non-responders (Lind & Anderson, 2019). We also inspected whether or not the clinical follow-up drug response observations of chemotherapeutic agents were the same with the classified prediction results.

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S1. Description of input datasets in the scenarios. Here, we set various combinations of dataset sizes, genomics data types, and drug information, and we assigned each combination to a dataset name (first column).

Scenarios (dataset names)	Genomics data types	Genomics vector length	Drug information	Drug chemical property vector length	Output (observed cell viability data)	Dataset size (training set)	Dataset size (test set)
Scenario 1 (MDG-160K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from GDSC	144,338	16,037
Scenario 2 (EDC-11K)	Expression (CCLE)	18,988	PaDEL descriptors	2,325	ln(IC50)s from CCLE	10,224	1,136
Scenario 3 (EYDC-9K)	Expression (CCLE)	18,988	PaDEL descriptors	2,325	ln(IC50)s from CCLE	8,631	959
	DNA methylation (CCLE)	20,192					
Scenario 4 (MDC-9K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from CCLE	7,854	873
Scenario 5 (ODC-11K)	Copy number variation (CCLE)	17,006	PaDEL descriptors	2,325	ln(IC50)s from CCLE	10,247	1,138
Scenario 6 (MEDG-61K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from GDSC	55,470	6,163
	Expression (CCLE)	18,988					
Scenario 7 (MEYDG-51K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from GDSC	46,756	5,195
	Expression (CCLE)	18,988					
	DNA methylation (CCLE)	20,192					
Scenario 8 (MEYODG-51K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from GDSC	46,609	5,178
	Expression (CCLE)	18,988					
	DNA methylation (CCLE)	20,192					
	Copy number variation (CCLE)	17,006					
Scenario 9 (MEDC-7K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from CCLE	6,436	715
	Expression (CCLE)	18,988					
Scenario 10 (MEYDC-6K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from CCLE	5,467	607
	Expression (CCLE)	18,988					
	DNA methylation (CCLE)	20,192					
Scenario 11 (MEYODC-6K)	Mutation (GDSC)	21,213	PaDEL descriptors	2,325	ln(IC50)s from CCLE	5,445	605
	Expression (CCLE)	18,988					
	DNA methylation (CCLE)	20,192					
	Copy number variation (CCLE)	17,006					

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S2. Model architecture and its parameters of CDRScan in scenarios 1 (MDG-160K) and 4 (MDC-9K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (Input-Layer)	(None, 21213, 1)	0	
drug_input (Input-Layer)	(None, 2325, 1)	0	
conv1d_5 (Conv1D)	(None, 4103, 50)	35050	cell_input[0][0]
conv1d_7 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 820, 50)	0	conv1d_5[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 141, 50)	0	conv1d_7[0][0]
conv1d_6 (Conv1D)	(None, 408, 30)	7530	max_pooling1d_5[0][0]
conv1d_8 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_7[0][0]
max_pooling1d_6 (MaxPooling1D)	(None, 40, 30)	0	conv1d_6[0][0]
max_pooling1d_8 (MaxPooling1D)	(None, 1, 30)	0	conv1d_8[0][0]
flatten_2 (Flatten)	(None, 1200)	0	max_pooling1d_6[0][0]
flatten_3 (Flatten)	(None, 30)	0	max_pooling1d_8[0][0]
dense_2 (Dense)	(None, 100)	120100	flatten_2[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dense_3 (Dense)	(None, 100)	3100	flatten_3[0][0]
dropout_4 (Dropout)	(None, 100)	0	dense_2[0][0]
dropout_5 (Dropout)	(None, 100)	0	dense_3[0][0]
concatenate (Concatenate)	(None, 200)	0	dropout_4[0][0] dropout_5[0][0]
dense_4 (Dense)	(None, 300)	60300	concatenate[0][0]
dropout_6 (Dropout)	(None, 300)	0	dense_4[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_6[0][0]
conv1d_9 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_9 (MaxPooling1D)	(None, 75, 30)	0	conv1d_9[0][0]
conv1d_10 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_9[0][0]
max_pooling1d_10 (MaxPooling1D)	(None, 23, 10)	0	conv1d_10[0][0]
conv1d_11 (Conv1D)	(None, 19, 5)	255	max_pooling1d_10[0][0]
max_pooling1d_11 (MaxPooling1D)	(None, 6, 5)	0	conv1d_11[0][0]
dropout_7 (Dropout)	(None, 6, 5)	0	max_pooling1d_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

flatten_4 (Flatten)	(None, 30)	0	dropout_7[0][0]
dropout_8 (Dropout)	(None, 30)	0	flatten_4[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_8[0][0]
Total params: 317,486 **Trainable params: 317,486** **Non-trainable params: 0**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S3. Model architecture and its parameters of CDRScan in scenario 2 (EDC-11K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (InputLayer)	(None, 18988, 1)	0	
drug_input (InputLayer)	(None, 2325, 1)	0	
conv1d_1 (Conv1D)	(None, 3658, 50)	35050	cell_input[0][0]
conv1d_3 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 731, 50)	0	conv1d_1[0][0]
max_pooling1d_3 (MaxPooling1D)	(None, 141, 50)	0	conv1d_3[0][0]
conv1d_2 (Conv1D)	(None, 364, 30)	7530	max_pooling1d_1[0][0]
conv1d_4 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_3[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 36, 30)	0	conv1d_2[0][0]
max_pooling1d_4 (MaxPooling1D)	(None, 1, 30)	0	conv1d_4[0][0]
flatten_1 (Flatten)	(None, 1080)	0	max_pooling1d_2[0][0]
flatten_2 (Flatten)	(None, 30)	0	max_pooling1d_4[0][0]
dense_1 (Dense)	(None, 100)	108100	flatten_1[0][0]
dense_2 (Dense)	(None, 100)	3100	flatten_2[0][0]
dropout_1 (Dropout)	(None, 100)	0	dense_1[0][0]
dropout_2 (Dropout)	(None, 100)	0	dense_2[0][0]
concatenate_1 (Concatenate)	(None, 200)	0	dropout_1[0][0] dropout_2[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dense_3 (Dense)	(None, 300)	60300	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 300)	0	dense_3[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_3[0][0]
conv1d_5 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 75, 30)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_5[0][0]
max_pooling1d_6 (MaxPooling1D)	(None, 23, 10)	0	conv1d_6[0][0]
conv1d_7 (Conv1D)	(None, 19, 5)	255	max_pooling1d_6[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 6, 5)	0	conv1d_7[0][0]
dropout_4 (Dropout)	(None, 6, 5)	0	max_pooling1d_7[0][0]
flatten_3 (Flatten)	(None, 30)	0	dropout_4[0][0]
dropout_5 (Dropout)	(None, 30)	0	flatten_3[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_5[0][0]
Total params: 305,486 **Trainable params: 305,486** **Non-trainable params: 0**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S4. Model architecture and its parameters of CDRScan in scenario 3 (EYDC-9K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (InputLayer)	(None, 39180, 1)	0	
drug_input (InputLayer)	(None, 2325, 1)	0	
conv1d_1 (Conv1D)	(None, 7697, 50)	35050	cell_input[0][0]
conv1d_3 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 1539, 50)	0	conv1d_1[0][0]
max_pooling1d_3 (MaxPooling1D)	(None, 141, 50)	0	conv1d_3[0][0]
conv1d_2 (Conv1D)	(None, 768, 30)	7530	max_pooling1d_1[0][0]
conv1d_4 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_3[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 76, 30)	0	conv1d_2[0][0]
max_pooling1d_4 (MaxPooling1D)	(None, 1, 30)	0	conv1d_4[0][0]
flatten_1 (Flatten)	(None, 2280)	0	max_pooling1d_2[0][0]
flatten_2 (Flatten)	(None, 30)	0	max_pooling1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dense_1 (Dense)	(None, 100)	228100	flatten_1[0][0]
dense_2 (Dense)	(None, 100)	3100	flatten_2[0][0]
dropout_1 (Dropout)	(None, 100)	0	dense_1[0][0]
dropout_2 (Dropout)	(None, 100)	0	dense_2[0][0]
concatenate_1 (Concatenate)	(None, 200)	0	dropout_1[0][0] dropout_2[0][0]
dense_3 (Dense)	(None, 300)	60300	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 300)	0	dense_3[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_3[0][0]
conv1d_5 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 75, 30)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_5[0][0]
max_pooling1d_6 (MaxPooling1D)	(None, 23, 10)	0	conv1d_6[0][0]
conv1d_7 (Conv1D)	(None, 19, 5)	255	max_pooling1d_6[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

max_pooling1d_7 (MaxPooling1D)	(None, 6, 5)	0	conv1d_7[0][0]
dropout_4 (Dropout)	(None, 6, 5)	0	max_pooling1d_7[0][0]
flatten_3 (Flatten)	(None, 30)	0	dropout_4[0][0]
dropout_5 (Dropout)	(None, 30)	0	flatten_3[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_5[0][0]
Total params: 425,486 **Trainable params: 425,486** **Non-trainable params: 0**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S5. Model architecture and its parameters of CDRScan in scenario 5 (ODC-11K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (Input-Layer)	(None, 17006, 1)	0	
drug_input (Input-Layer)	(None, 2325, 1)	0	
conv1d_5 (Conv1D)	(None, 3262, 50)	35050	cell_input[0][0]
conv1d_7 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 652, 50)	0	conv1d_5[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 141, 50)	0	conv1d_7[0][0]
conv1d_6 (Conv1D)	(None, 324, 30)	7530	max_pooling1d_5[0][0]
conv1d_8 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_7[0][0]
max_pooling1d_6 (MaxPooling1D)	(None, 32, 30)	0	conv1d_6[0][0]
max_pooling1d_8 (MaxPooling1D)	(None, 1, 30)	0	conv1d_8[0][0]
flatten_2 (Flatten)	(None, 960)	0	max_pooling1d_6[0][0]
flatten_3 (Flatten)	(None, 30)	0	max_pooling1d_8[0][0]
dense_2 (Dense)	(None, 100)	96100	flatten_2[0][0]
dense_3 (Dense)	(None, 100)	3100	flatten_3[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dropout_4 (Dropout)	(None, 100)	0	dense_2[0][0]
dropout_5 (Dropout)	(None, 100)	0	dense_3[0][0]
concatenate (Concatenate)	(None, 200)	0	dropout_4[0][0] dropout_5[0][0]
dense_4 (Dense)	(None, 300)	60300	concatenate[0][0]
dropout_6 (Dropout)	(None, 300)	0	dense_4[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_6[0][0]
conv1d_9 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_9 (MaxPooling1D)	(None, 75, 30)	0	conv1d_9[0][0]
conv1d_10 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_9[0][0]
max_pooling1d_10 (MaxPooling1D)	(None, 23, 10)	0	conv1d_10[0][0]
conv1d_11 (Conv1D)	(None, 19, 5)	255	max_pooling1d_10[0][0]
max_pooling1d_11 (MaxPooling1D)	(None, 6, 5)	0	conv1d_11[0][0]
dropout_7 (Dropout)	(None, 6, 5)	0	max_pooling1d_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

flatten_4 (Flatten)	(None, 30)	0	dropout_7[0][0]
dropout_8 (Dropout)	(None, 30)	0	flatten_4[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_8[0][0]
Total params: 293,486 **Trainable params: 293,486** **Non-trainable params: 0**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S6. Model architecture and its parameters of CDRScan in scenarios 6 (MEDG-61K) and 9 (MEDC-7K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (InputLayer)	(None, 40201, 1)	0	
drug_input (InputLayer)	(None, 2325, 1)	0	
conv1d_1 (Conv1D)	(None, 7901, 50)	35050	cell_input[0][0]
conv1d_3 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 1580, 50)	0	conv1d_1[0][0]
max_pooling1d_3 (MaxPooling1D)	(None, 141, 50)	0	conv1d_3[0][0]
conv1d_2 (Conv1D)	(None, 788, 30)	7530	max_pooling1d_1[0][0]
conv1d_4 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_3[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 78, 30)	0	conv1d_2[0][0]
max_pooling1d_4 (MaxPooling1D)	(None, 1, 30)	0	conv1d_4[0][0]
flatten_1 (Flatten)	(None, 2340)	0	max_pooling1d_2[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

flatten_2 (Flatten)	(None, 30)	0	max_pooling1d_4[0][0]
dense_1 (Dense)	(None, 100)	234100	flatten_1[0][0]
dense_2 (Dense)	(None, 100)	3100	flatten_2[0][0]
dropout_1 (Dropout)	(None, 100)	0	dense_1[0][0]
dropout_2 (Dropout)	(None, 100)	0	dense_2[0][0]
concatenate_1 (Concatenate)	(None, 200)	0	dropout_1[0][0] dropout_2[0][0]
dense_3 (Dense)	(None, 300)	60300	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 300)	0	dense_3[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_3[0][0]
conv1d_5 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 75, 30)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

max_pooling1d_6 (MaxPooling1D)	(None, 23, 10)	0	conv1d_6[0][0]
conv1d_7 (Conv1D)	(None, 19, 5)	255	max_pooling1d_6[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 6, 5)	0	conv1d_7[0][0]
dropout_4 (Dropout)	(None, 6, 5)	0	max_pooling1d_7[0][0]
flatten_3 (Flatten)	(None, 30)	0	dropout_4[0][0]
dropout_5 (Dropout)	(None, 30)	0	flatten_3[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_5[0][0]
Total params: 431,486 Trainable params: 431,486 Non-trainable params: 0			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S7. Model architecture and its parameters of CDRScan in scenarios 7 (MEYDG-51K) and 10 (MEYDC-6K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (InputLayer)	(None, 60393, 1)	0	
drug_input (InputLayer)	(None, 2325, 1)	0	
conv1d_1 (Conv1D)	(None, 11939, 50)	35050	cell_input[0][0]
conv1d_3 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 2387, 50)	0	conv1d_1[0][0]
max_pooling1d_3 (MaxPooling1D)	(None, 141, 50)	0	conv1d_3[0][0]
conv1d_2 (Conv1D)	(None, 1192, 30)	7530	max_pooling1d_1[0][0]
conv1d_4 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_3[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 119, 30)	0	conv1d_2[0][0]
max_pooling1d_4 (MaxPooling1D)	(None, 1, 30)	0	conv1d_4[0][0]
flatten_1 (Flatten)	(None, 3570)	0	max_pooling1d_2[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

flatten_2 (Flatten)	(None, 30)	0	max_pooling1d_4[0][0]
dense_1 (Dense)	(None, 100)	357100	flatten_1[0][0]
dense_2 (Dense)	(None, 100)	3100	flatten_2[0][0]
dropout_1 (Dropout)	(None, 100)	0	dense_1[0][0]
dropout_2 (Dropout)	(None, 100)	0	dense_2[0][0]
concatenate_1 (Concatenate)	(None, 200)	0	dropout_1[0][0] dropout_2[0][0]
dense_3 (Dense)	(None, 300)	60300	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 300)	0	dense_3[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_3[0][0]
conv1d_5 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 75, 30)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

max_pooling1d_6 (MaxPooling1D)	(None, 23, 10)	0	conv1d_6[0][0]
conv1d_7 (Conv1D)	(None, 19, 5)	255	max_pooling1d_6[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 6, 5)	0	conv1d_7[0][0]
dropout_4 (Dropout)	(None, 6, 5)	0	max_pooling1d_7[0][0]
flatten_3 (Flatten)	(None, 30)	0	dropout_4[0][0]
dropout_5 (Dropout)	(None, 30)	0	flatten_3[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_5[0][0]
Total params: 554,486 Trainable params: 554,486 Non-trainable params: 0			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S8. Model architecture and its parameters of CDRScan in scenarios 9 (MEYODG-51K) and 11 (MEYODC-6K).

Layer (type)	Output Shape	Param #	Connected to
cell_input (InputLayer)	(None, 77399, 1)	0	
drug_input (InputLayer)	(None, 2325, 1)	0	
conv1d_1 (Conv1D)	(None, 15340, 50)	35050	cell_input[0][0]
conv1d_3 (Conv1D)	(None, 709, 50)	10050	drug_input[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 3068, 50)	0	conv1d_1[0][0]
max_pooling1d_3 (MaxPooling1D)	(None, 141, 50)	0	conv1d_3[0][0]
conv1d_2 (Conv1D)	(None, 1532, 30)	7530	max_pooling1d_1[0][0]
conv1d_4 (Conv1D)	(None, 19, 30)	75030	max_pooling1d_3[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 153, 30)	0	conv1d_2[0][0]
max_pooling1d_4 (MaxPooling1D)	(None, 1, 30)	0	conv1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

flatten_1 (Flatten)	(None, 4590)	0	max_pooling1d_2[0][0]
flatten_2 (Flatten)	(None, 30)	0	max_pooling1d_4[0][0]
dense_1 (Dense)	(None, 100)	459100	flatten_1[0][0]
dense_2 (Dense)	(None, 100)	3100	flatten_2[0][0]
dropout_1 (Dropout)	(None, 100)	0	dense_1[0][0]
dropout_2 (Dropout)	(None, 100)	0	dense_2[0][0]
concatenate_1 (Concatenate)	(None, 200)	0	dropout_1[0][0] dropout_2[0][0]
dense_3 (Dense)	(None, 300)	60300	concatenate_1[0][0]
dropout_3 (Dropout)	(None, 300)	0	dense_3[0][0]
reshape_1 (Reshape)	(None, 300, 1)	0	dropout_3[0][0]
conv1d_5 (Conv1D)	(None, 151, 30)	4530	reshape_1[0][0]
max_pooling1d_5 (MaxPooling1D)	(None, 75, 30)	0	conv1d_5[0][0]
conv1d_6 (Conv1D)	(None, 71, 10)	1510	max_pooling1d_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

max_pooling1d_6 (MaxPooling1D)	(None, 23, 10)	0	conv1d_6[0][0]
conv1d_7 (Conv1D)	(None, 19, 5)	255	max_pooling1d_6[0][0]
max_pooling1d_7 (MaxPooling1D)	(None, 6, 5)	0	conv1d_7[0][0]
dropout_4 (Dropout)	(None, 6, 5)	0	max_pooling1d_7[0][0]
flatten_3 (Flatten)	(None, 30)	0	dropout_4[0][0]
dropout_5 (Dropout)	(None, 30)	0	flatten_3[0][0]
pred_lnIC50 (Dense)	(None, 1)	31	dropout_5[0][0]
Total params: 656,486 Trainable params: 656,486 Non-trainable params: 0			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S9. Parameters of CDRScan models in all scenarios.

Scenarios	Total parameters	Total number of layers	The number of convolution Layers		The number of fully connected layers	Architecture	Loss function	Optimizer	Learning rate	Training epoch	Batch size
			For genomics data	For drug data							
Scenario 1 (MDG-160K)	317,486	18	5	5	8	Supplementary Table S2 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 2 (EDC-11K)	305,486	18	5	5	8	Supplementary Table S3 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 3 (EYDC-9K)	425,486	18	5	5	8	Supplementary Table S4 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 4 (MDC-9K)	317,486	18	5	5	8	Supplementary Table S2 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 5 (ODC-11K)	293,486	18	5	5	8	Supplementary Table S5 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 6 (MEDG-61K)	431,486	18	5	5	8	Supplementary Table S6 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 7 (MEYDG-51K)	554,486	18	5	5	8	Supplementary Table S7 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 8 (MEYODG-51K)	656,486	18	5	5	8	Supplementary Table S8 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 9 (MEDC-7K)	431,486	18	5	5	8	Supplementary Table S6 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 10 (MEYDC-6K)	554,486	18	5	5	8	Supplementary Table S7 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 11 (MEYODC-6K)	656,486	18	5	5	8	Supplementary Table S8 and Supplementary Figure S2	Mean square error (MSE)	Adam	0.0002	150	100

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S10. Model architecture and its parameters of ResNet in scenarios 1 (MDG-160K) and 4 (MDC-9K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 23538, 1)	0	
conv1d_1 (Conv1D)	(None, 11769, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 2353, 16)	0	conv1d_1[0][0]
batch_normalization_1 (Batch_normalization)	(None, 2353, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 2353, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 2353, 16)	784	activation_1[0][0]
batch_normalization_2 (Batch_normalization)	(None, 2353, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 2353, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 2353, 16)	784	activation_2[0][0]
batch_normalization_3 (Batch_normalization)	(None, 2353, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 2353, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 2353, 16)	784	activation_3[0][0]
batch_normalization_4 (Batch_normalization)	(None, 2353, 16)	64	conv1d_4[0][0]
add_1 (Add)	(None, 2353, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_4 (Activation)	(None, 2353, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 2353, 16)	784	activation_4[0][0]
batch_normalization_5 (Batch_normalization)	(None, 2353, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 2353, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 2353, 16)	784	activation_5[0][0]
batch_normalization_6 (Batch_normalization)	(None, 2353, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 2353, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 2353, 16)	784	activation_6[0][0]
batch_normalization_7 (Batch_normalization)	(None, 2353, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 2353, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 2353, 16)	784	activation_7[0][0]
batch_normalization_8 (Batch_normalization)	(None, 2353, 16)	64	conv1d_8[0][0]
add_2 (Add)	(None, 2353, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]
activation_8 (Activation)	(None, 2353, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 1177, 32)	1568	activation_8[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_9 (Batch_normalization)	(None, 1177, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 1177, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 1177, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 1177, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (Batch_normalization)	(None, 1177, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 1177, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 1177, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 1177, 32)	3104	activation_10[0][0]
batch_normalization_11 (Batch_normalization)	(None, 1177, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 1177, 32)	0	batch_normalization_11[0][0]
conv1d_13 (Conv1D)	(None, 1177, 32)	3104	activation_11[0][0]
batch_normalization_12 (Batch_normalization)	(None, 1177, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 1177, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 1177, 32)	0	add_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_14 (Conv1D)	(None, 1177, 32)	3104	activation_12[0][0]
batch_normalization_13 (Batch_normalization)	(None, 1177, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 1177, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 1177, 32)	3104	activation_13[0][0]
batch_normalization_14 (Batch_normalization)	(None, 1177, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 1177, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 1177, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 589, 64)	6208	activation_14[0][0]
batch_normalization_15 (Batch_normalization)	(None, 589, 64)	256	conv1d_16[0][0]
activation_15 (Activation)	(None, 589, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 589, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 589, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (Batch_normalization)	(None, 589, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 589, 64)	0	batch_normalization_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

			batch_normalization_15[0][0]
activation_16 (Activation)	(None, 589, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 589, 64)	12352	activation_16[0][0]
batch_normalization_17 (Batch_normalization)	(None, 589, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 589, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 589, 64)	12352	activation_17[0][0]
batch_normalization_18 (Batch_normalization)	(None, 589, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 589, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 589, 64)	0	add_7[0][0]
conv1d_21 (Conv1D)	(None, 589, 64)	12352	activation_18[0][0]
batch_normalization_19 (Batch_normalization)	(None, 589, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 589, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 589, 64)	12352	activation_19[0][0]
batch_normalization_20 (Batch_normalization)	(None, 589, 64)	256	conv1d_22[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_8 (Add)	(None, 589, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 589, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 37696)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	77203456	flatten_1[0][0]
batch_normalization_21 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]
batch_normalization_22 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_23[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]
batch_normalization_25 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_27 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 81,171,793 **Trainable params: 81,159,377** **Non-trainable params: 12,416**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S11. Model architecture and its parameters of ResNet in scenario 2 (EDC-11K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 21313, 1)	0	
conv1d_23 (Conv1D)	(None, 10657, 16)	64	inputs[0][0]
max_pooling1d_2 (MaxPooling1D)	(None, 2131, 16)	0	conv1d_23[0][0]
batch_normalization_28 (Batch_normalization)	(None, 2131, 16)	64	max_pooling1d_2[0][0]
activation_28 (Activation)	(None, 2131, 16)	0	batch_normalization_28[0][0]
conv1d_24 (Conv1D)	(None, 2131, 16)	784	activation_28[0][0]
batch_normalization_29 (Batch_normalization)	(None, 2131, 16)	64	conv1d_24[0][0]
activation_29 (Activation)	(None, 2131, 16)	0	batch_normalization_29[0][0]
conv1d_25 (Conv1D)	(None, 2131, 16)	784	activation_29[0][0]
batch_normalization_30 (Batch_normalization)	(None, 2131, 16)	64	conv1d_25[0][0]
activation_30 (Activation)	(None, 2131, 16)	0	batch_normalization_30[0][0]
conv1d_26 (Conv1D)	(None, 2131, 16)	784	activation_30[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_31 (Batch_normalization)	(None, 2131, 16)	64	conv1d_26[0][0]
add_10 (Add)	(None, 2131, 16)	0	batch_normalization_31[0][0] batch_normalization_29[0][0]
activation_31 (Activation)	(None, 2131, 16)	0	add_10[0][0]
conv1d_27 (Conv1D)	(None, 2131, 16)	784	activation_31[0][0]
batch_normalization_32 (Batch_normalization)	(None, 2131, 16)	64	conv1d_27[0][0]
activation_32 (Activation)	(None, 2131, 16)	0	batch_normalization_32[0][0]
conv1d_28 (Conv1D)	(None, 2131, 16)	784	activation_32[0][0]
batch_normalization_33 (Batch_normalization)	(None, 2131, 16)	64	conv1d_28[0][0]
activation_33 (Activation)	(None, 2131, 16)	0	batch_normalization_33[0][0]
conv1d_29 (Conv1D)	(None, 2131, 16)	784	activation_33[0][0]
batch_normalization_34 (Batch_normalization)	(None, 2131, 16)	64	conv1d_29[0][0]
activation_34 (Activation)	(None, 2131, 16)	0	batch_normalization_34[0][0]
conv1d_30 (Conv1D)	(None, 2131, 16)	784	activation_34[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_35 (Batch_normalization)	(None, 2131, 16)	64	conv1d_30[0][0]
add_11 (Add)	(None, 2131, 16)	0	batch_normalization_35[0][0] batch_normalization_32[0][0]
activation_35 (Activation)	(None, 2131, 16)	0	add_11[0][0]
conv1d_31 (Conv1D)	(None, 1066, 32)	1568	activation_35[0][0]
batch_normalization_36 (Batch_normalization)	(None, 1066, 32)	128	conv1d_31[0][0]
activation_36 (Activation)	(None, 1066, 32)	0	batch_normalization_36[0][0]
conv1d_32 (Conv1D)	(None, 1066, 32)	3104	activation_36[0][0]
conv1d_33 (Conv1D)	(None, 1066, 32)	3104	conv1d_32[0][0]
batch_normalization_37 (Batch_normalization)	(None, 1066, 32)	128	conv1d_33[0][0]
add_12 (Add)	(None, 1066, 32)	0	batch_normalization_37[0][0] batch_normalization_36[0][0]
activation_37 (Activation)	(None, 1066, 32)	0	add_12[0][0]
conv1d_34 (Conv1D)	(None, 1066, 32)	3104	activation_37[0][0]
batch_normalization_38 (Batch_normalization)	(None, 1066, 32)	128	conv1d_34[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_38 (Activation)	(None, 1066, 32)	0	batch_normalization_38[0][0]
conv1d_35 (Conv1D)	(None, 1066, 32)	3104	activation_38[0][0]
batch_normalization_39 (Batch_normalization)	(None, 1066, 32)	128	conv1d_35[0][0]
add_13 (Add)	(None, 1066, 32)	0	batch_normalization_39[0][0] batch_normalization_38[0][0]
activation_39 (Activation)	(None, 1066, 32)	0	add_13[0][0]
conv1d_36 (Conv1D)	(None, 1066, 32)	3104	activation_39[0][0]
batch_normalization_40 (Batch_normalization)	(None, 1066, 32)	128	conv1d_36[0][0]
activation_40 (Activation)	(None, 1066, 32)	0	batch_normalization_40[0][0]
conv1d_37 (Conv1D)	(None, 1066, 32)	3104	activation_40[0][0]
batch_normalization_41 (Batch_normalization)	(None, 1066, 32)	128	conv1d_37[0][0]
add_14 (Add)	(None, 1066, 32)	0	batch_normalization_41[0][0] batch_normalization_40[0][0]
activation_41 (Activation)	(None, 1066, 32)	0	add_14[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_38 (Conv1D)	(None, 533, 64)	6208	activation_41[0][0]
batch_normalization_42 (Batch_normalization)	(None, 533, 64)	256	conv1d_38[0][0]
activation_42 (Activation)	(None, 533, 64)	0	batch_normalization_42[0][0]
conv1d_39 (Conv1D)	(None, 533, 64)	12352	activation_42[0][0]
conv1d_40 (Conv1D)	(None, 533, 64)	12352	conv1d_39[0][0]
batch_normalization_43 (Batch_normalization)	(None, 533, 64)	256	conv1d_40[0][0]
add_15 (Add)	(None, 533, 64)	0	batch_normalization_43[0][0] batch_normalization_42[0][0]
activation_43 (Activation)	(None, 533, 64)	0	add_15[0][0]
conv1d_41 (Conv1D)	(None, 533, 64)	12352	activation_43[0][0]
batch_normalization_44 (Batch_normalization)	(None, 533, 64)	256	conv1d_41[0][0]
activation_44 (Activation)	(None, 533, 64)	0	batch_normalization_44[0][0]
conv1d_42 (Conv1D)	(None, 533, 64)	12352	activation_44[0][0]
batch_normalization_45 (Batch_normalization)	(None, 533, 64)	256	conv1d_42[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_16 (Add)	(None, 533, 64)	0	batch_normalization_45[0][0] batch_normalization_44[0][0]
activation_45 (Activation)	(None, 533, 64)	0	add_16[0][0]
conv1d_43 (Conv1D)	(None, 533, 64)	12352	activation_45[0][0]
batch_normalization_46 (Batch_normalization)	(None, 533, 64)	256	conv1d_43[0][0]
activation_46 (Activation)	(None, 533, 64)	0	batch_normalization_46[0][0]
conv1d_44 (Conv1D)	(None, 533, 64)	12352	activation_46[0][0]
batch_normalization_47 (Batch_normalization)	(None, 533, 64)	256	conv1d_44[0][0]
add_17 (Add)	(None, 533, 64)	0	batch_normalization_47[0][0] batch_normalization_46[0][0]
activation_47 (Activation)	(None, 533, 64)	0	add_17[0][0]
flatten_2 (Flatten)	(None, 34112)	0	activation_47[0][0]
dense1 (Dense)	(None, 2048)	69863424	flatten_2[0][0]
batch_normalization_48 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_48[0][0]
activation_48 (Activation)	(None, 2048)	0	dropout1[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dense5 (Dense)	(None, 1024)	2098176	activation_48[0][0]
batch_normalization_49 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_49[0][0]
activation_49 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_49[0][0]
batch_normalization_50 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_50[0][0]
activation_50 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_50[0][0]
batch_normalization_51 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_51[0][0]
add_18 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_49[0][0]
activation_51 (Activation)	(None, 1024)	0	add_18[0][0]
dense8 (Dense)	(None, 512)	524800	activation_51[0][0]
batch_normalization_52 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_52[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_52 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_52[0][0]
batch_normalization_53 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_53[0][0]
activation_53 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_53[0][0]
batch_normalization_54 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_54[0][0]
activation_54 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_54[0][0]
Total params: 73,831,761 **Trainable params: 73,819,345** **Non-trainable params: 12,416**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S12. Model architecture and its parameters of ResNet in scenario 3 (EYDC-9K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 41505, 1)	0	
conv1d_1 (Conv1D)	(None, 20753, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 4150, 16)	0	conv1d_1[0][0]
batch_normalization_1 (BatchNormalization)	(None, 4150, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 4150, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 4150, 16)	784	activation_1[0][0]
batch_normalization_2 (BatchNormalization)	(None, 4150, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 4150, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 4150, 16)	784	activation_2[0][0]
batch_normalization_3 (BatchNormalization)	(None, 4150, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 4150, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 4150, 16)	784	activation_3[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_4 (BatchNormalization)	(None, 4150, 16)	64	conv1d_4[0][0]
add_1 (Add)	(None, 4150, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]
activation_4 (Activation)	(None, 4150, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 4150, 16)	784	activation_4[0][0]
batch_normalization_5 (BatchNormalization)	(None, 4150, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 4150, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 4150, 16)	784	activation_5[0][0]
batch_normalization_6 (BatchNormalization)	(None, 4150, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 4150, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 4150, 16)	784	activation_6[0][0]
batch_normalization_7 (BatchNormalization)	(None, 4150, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 4150, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 4150, 16)	784	activation_7[0][0]
batch_normalization_8 (BatchNormalization)	(None, 4150, 16)	64	conv1d_8[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_2 (Add)	(None, 4150, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]
activation_8 (Activation)	(None, 4150, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 2075, 32)	1568	activation_8[0][0]
batch_normalization_9 (BatchNormalization)	(None, 2075, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 2075, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 2075, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 2075, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (BatchNormalization)	(None, 2075, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 2075, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 2075, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 2075, 32)	3104	activation_10[0][0]
batch_normalization_11 (BatchNormalization)	(None, 2075, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 2075, 32)	0	batch_normalization_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_13 (Conv1D)	(None, 2075, 32)	3104	activation_11[0][0]
batch_normalization_12 (BatchNormalization)	(None, 2075, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 2075, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 2075, 32)	0	add_4[0][0]
conv1d_14 (Conv1D)	(None, 2075, 32)	3104	activation_12[0][0]
batch_normalization_13 (BatchNormalization)	(None, 2075, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 2075, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 2075, 32)	3104	activation_13[0][0]
batch_normalization_14 (BatchNormalization)	(None, 2075, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 2075, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 2075, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 1038, 64)	6208	activation_14[0][0]
batch_normalization_15 (BatchNormalization)	(None, 1038, 64)	256	conv1d_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_15 (Activation)	(None, 1038, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 1038, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 1038, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (BatchNormalization)	(None, 1038, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 1038, 64)	0	batch_normalization_16[0][0] batch_normalization_15[0][0]
activation_16 (Activation)	(None, 1038, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 1038, 64)	12352	activation_16[0][0]
batch_normalization_17 (BatchNormalization)	(None, 1038, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 1038, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 1038, 64)	12352	activation_17[0][0]
batch_normalization_18 (BatchNormalization)	(None, 1038, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 1038, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 1038, 64)	0	add_7[0][0]
conv1d_21 (Conv1D)	(None, 1038, 64)	12352	activation_18[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_19 (BatchNormalization)	(None, 1038, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 1038, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 1038, 64)	12352	activation_19[0][0]
batch_normalization_20 (BatchNormalization)	(None, 1038, 64)	256	conv1d_22[0][0]
add_8 (Add)	(None, 1038, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 1038, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 66432)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	136054784	flatten_1[0][0]
batch_normalization_21 (BatchNormalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]
batch_normalization_22 (BatchNormalization)	(None, 1024)	4096	dense5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dropout5 (Dropout)	(None, 1024)	0	batch_normaliza-tion_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (BatchNormalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normaliza-tion_23[0][0]
activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (BatchNormalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normaliza-tion_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normaliza-tion_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]
batch_normalization_25 (BatchNormalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normaliza-tion_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (BatchNormalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normaliza-tion_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]
batch_normalization_27 (BatchNormalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normaliza-tion_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 140,023,121 **Trainable params: 140,010,705** **Non-trainable params: 12,416**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S13. Model architecture and its parameters of ResNet in scenario 5 (ODC-11K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 19331, 1)	0	
conv1d_1 (Conv1D)	(None, 9666, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 1933, 16)	0	conv1d_1[0][0]
batch_normalization_1 (Batch_normalization)	(None, 1933, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 1933, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 1933, 16)	784	activation_1[0][0]
batch_normalization_2 (Batch_normalization)	(None, 1933, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 1933, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 1933, 16)	784	activation_2[0][0]
batch_normalization_3 (Batch_normalization)	(None, 1933, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 1933, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 1933, 16)	784	activation_3[0][0]
batch_normalization_4 (Batch_normalization)	(None, 1933, 16)	64	conv1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_1 (Add)	(None, 1933, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]
activation_4 (Activation)	(None, 1933, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 1933, 16)	784	activation_4[0][0]
batch_normalization_5 (Batch_normalization)	(None, 1933, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 1933, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 1933, 16)	784	activation_5[0][0]
batch_normalization_6 (Batch_normalization)	(None, 1933, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 1933, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 1933, 16)	784	activation_6[0][0]
batch_normalization_7 (Batch_normalization)	(None, 1933, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 1933, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 1933, 16)	784	activation_7[0][0]
batch_normalization_8 (Batch_normalization)	(None, 1933, 16)	64	conv1d_8[0][0]
add_2 (Add)	(None, 1933, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_8 (Activation)	(None, 1933, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 967, 32)	1568	activation_8[0][0]
batch_normalization_9 (Batch_normalization)	(None, 967, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 967, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 967, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 967, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (Batch_normalization)	(None, 967, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 967, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 967, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 967, 32)	3104	activation_10[0][0]
batch_normalization_11 (Batch_normalization)	(None, 967, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 967, 32)	0	batch_normalization_11[0][0]
conv1d_13 (Conv1D)	(None, 967, 32)	3104	activation_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_12 (Batch_normalization)	(None, 967, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 967, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 967, 32)	0	add_4[0][0]
conv1d_14 (Conv1D)	(None, 967, 32)	3104	activation_12[0][0]
batch_normalization_13 (Batch_normalization)	(None, 967, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 967, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 967, 32)	3104	activation_13[0][0]
batch_normalization_14 (Batch_normalization)	(None, 967, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 967, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 967, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 484, 64)	6208	activation_14[0][0]
batch_normalization_15 (Batch_normalization)	(None, 484, 64)	256	conv1d_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_15 (Activation)	(None, 484, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 484, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 484, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (Batch_normalization)	(None, 484, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 484, 64)	0	batch_normalization_16[0][0] batch_normalization_15[0][0]
activation_16 (Activation)	(None, 484, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 484, 64)	12352	activation_16[0][0]
batch_normalization_17 (Batch_normalization)	(None, 484, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 484, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 484, 64)	12352	activation_17[0][0]
batch_normalization_18 (Batch_normalization)	(None, 484, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 484, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 484, 64)	0	add_7[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_21 (Conv1D)	(None, 484, 64)	12352	activation_18[0][0]
batch_normalization_19 (Batch_normalization)	(None, 484, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 484, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 589, 64)	12352	activation_19[0][0]
batch_normalization_20 (Batch_normalization)	(None, 484, 64)	256	conv1d_22[0][0]
add_8 (Add)	(None, 484, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 484, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 30976)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	63440896	flatten_1[0][0]
batch_normalization_21 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_22 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_23[0][0]
activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_25 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]
batch_normalization_27 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 67,409,233 **Trainable params: 67,396,817**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Non-trainable params: 12,416			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S14. Model architecture and its parameters of ResNet in scenarios 6 (MEDG-61K) and 9 (MEDC-6K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 42526, 1)	0	
conv1d_1 (Conv1D)	(None, 21216, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 4252, 16)	0	conv1d_1[0][0]
batch_normalization_1 (Batch_normalization)	(None, 4252, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 4252, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 4252, 16)	784	activation_1[0][0]
batch_normalization_2 (Batch_normalization)	(None, 4252, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 4252, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 4252, 16)	784	activation_2[0][0]
batch_normalization_3 (Batch_normalization)	(None, 4252, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 4252, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 4252, 16)	784	activation_3[0][0]
batch_normalization_4 (Batch_normalization)	(None, 4252, 16)	64	conv1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_1 (Add)	(None, 4252, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]
activation_4 (Activation)	(None, 4252, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 4252, 16)	784	activation_4[0][0]
batch_normalization_5 (Batch_normalization)	(None, 4252, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 4252, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 4252, 16)	784	activation_5[0][0]
batch_normalization_6 (Batch_normalization)	(None, 4252, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 4252, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 4252, 16)	784	activation_6[0][0]
batch_normalization_7 (Batch_normalization)	(None, 4252, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 4252, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 4252, 16)	784	activation_7[0][0]
batch_normalization_8 (Batch_normalization)	(None, 4252, 16)	64	conv1d_8[0][0]
add_2 (Add)	(None, 4252, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_8 (Activation)	(None, 4252, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 2126, 32)	1568	activation_8[0][0]
batch_normalization_9 (Batch_normalization)	(None, 2126, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 2126, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 2126, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 2126, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (Batch_normalization)	(None, 2126, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 2126, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 2126, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 2126, 32)	3104	activation_10[0][0]
batch_normalization_11 (Batch_normalization)	(None, 2126, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 2126, 32)	0	batch_normalization_11[0][0]
conv1d_13 (Conv1D)	(None, 2126, 32)	3104	activation_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_12 (Batch_normalization)	(None, 2126, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 2126, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 2126, 32)	0	add_4[0][0]
conv1d_14 (Conv1D)	(None, 2126, 32)	3104	activation_12[0][0]
batch_normalization_13 (Batch_normalization)	(None, 2126, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 2126, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 2126, 32)	3104	activation_13[0][0]
batch_normalization_14 (Batch_normalization)	(None, 2126, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 2126, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 2126, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 1063, 64)	6208	activation_14[0][0]
batch_normalization_15 (Batch_normalization)	(None, 1063, 64)	256	conv1d_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_15 (Activation)	(None, 1063, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 1063, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 1063, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (Batch_normalization)	(None, 1063, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 1063, 64)	0	batch_normalization_16[0][0] batch_normalization_15[0][0]
activation_16 (Activation)	(None, 1063, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 1063, 64)	12352	activation_16[0][0]
batch_normalization_17 (Batch_normalization)	(None, 1063, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 1063, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 1063, 64)	12352	activation_17[0][0]
batch_normalization_18 (Batch_normalization)	(None, 1063, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 1063, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 1063, 64)	0	add_7[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_21 (Conv1D)	(None, 1063, 64)	12352	activation_18[0][0]
batch_normalization_19 (Batch_normalization)	(None, 1063, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 1063, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 1063, 64)	12352	activation_19[0][0]
batch_normalization_20 (Batch_normalization)	(None, 1063, 64)	256	conv1d_22[0][0]
add_8 (Add)	(None, 1063, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 1063, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 68032)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	139331584	flatten_1[0][0]
batch_normalization_21 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_22 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_23[0][0]
activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_25 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]
batch_normalization_27 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 143,299,921 **Trainable params: 143,287,505**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Non-trainable params: 12,416			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S15. Model architecture and its parameters of ResNet in scenarios 7 (MEYDG-51K) and 10 (MEYDC-6K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 62718, 1)	0	
conv1d_1 (Conv1D)	(None, 31359, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 6271, 16)	0	conv1d_1[0][0]
batch_normalization_1 (Batch_normalization)	(None, 6271, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 6271, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 6271, 16)	784	activation_1[0][0]
batch_normalization_2 (Batch_normalization)	(None, 6271, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 6271, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 6271, 16)	784	activation_2[0][0]
batch_normalization_3 (Batch_normalization)	(None, 6271, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 6271, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 6271, 16)	784	activation_3[0][0]
batch_normalization_4 (Batch_normalization)	(None, 6271, 16)	64	conv1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_1 (Add)	(None, 6271, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]
activation_4 (Activation)	(None, 6271, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 6271, 16)	784	activation_4[0][0]
batch_normalization_5 (Batch_normalization)	(None, 6271, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 6271, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 6271, 16)	784	activation_5[0][0]
batch_normalization_6 (Batch_normalization)	(None, 6271, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 6271, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 6271, 16)	784	activation_6[0][0]
batch_normalization_7 (Batch_normalization)	(None, 6271, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 6271, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 6271, 16)	784	activation_7[0][0]
batch_normalization_8 (Batch_normalization)	(None, 6271, 16)	64	conv1d_8[0][0]
add_2 (Add)	(None, 6271, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_8 (Activation)	(None, 6271, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 3136, 32)	1568	activation_8[0][0]
batch_normalization_9 (Batch_normalization)	(None, 3136, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 3136, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 3136, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 3136, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (Batch_normalization)	(None, 3136, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 3136, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 3136, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 3136, 32)	3104	activation_10[0][0]
batch_normalization_11 (Batch_normalization)	(None, 3136, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 3136, 32)	0	batch_normalization_11[0][0]
conv1d_13 (Conv1D)	(None, 3136, 32)	3104	activation_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_12 (Batch_normalization)	(None, 3136, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 3136, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 3136, 32)	0	add_4[0][0]
conv1d_14 (Conv1D)	(None, 3136, 32)	3104	activation_12[0][0]
batch_normalization_13 (Batch_normalization)	(None, 3136, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 3136, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 3136, 32)	3104	activation_13[0][0]
batch_normalization_14 (Batch_normalization)	(None, 3136, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 3136, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 3136, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 1568, 64)	6208	activation_14[0][0]
batch_normalization_15 (Batch_normalization)	(None, 1568, 64)	256	conv1d_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_15 (Activation)	(None, 1568, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 1568, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 1568, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (Batch_normalization)	(None, 1568, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 1568, 64)	0	batch_normalization_16[0][0] batch_normalization_15[0][0]
activation_16 (Activation)	(None, 1568, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 1568, 64)	12352	activation_16[0][0]
batch_normalization_17 (Batch_normalization)	(None, 1568, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 1568, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 1568, 64)	12352	activation_17[0][0]
batch_normalization_18 (Batch_normalization)	(None, 1568, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 1568, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 1568, 64)	0	add_7[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_21 (Conv1D)	(None, 1568, 64)	12352	activation_18[0][0]
batch_normalization_19 (Batch_normalization)	(None, 1568, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 1568, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 1568, 64)	12352	activation_19[0][0]
batch_normalization_20 (Batch_normalization)	(None, 1568, 64)	256	conv1d_22[0][0]
add_8 (Add)	(None, 1568, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 1063, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 100352)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	205522944	flatten_1[0][0]
batch_normalization_21 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_22 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_23[0][0]
activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_25 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]
batch_normalization_27 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 209,491,281 Trainable params: 209,478,865			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Non-trainable params: 12,416			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S16. Model architecture and its parameters of ResNet in scenarios 8 (MEYODG-51K) and 11 (MEYODC-6K).

Layer (type)	Output Shape	Param #	Connected to
inputs (InputLayer)	(None, 79724, 1)	0	
conv1d_1 (Conv1D)	(None, 39862, 16)	64	inputs[0][0]
max_pooling1d_1 (MaxPooling1D)	(None, 7972, 16)	0	conv1d_1[0][0]
batch_normalization_1 (Batch_normalization)	(None, 7972, 16)	64	max_pooling1d_1[0][0]
activation_1 (Activation)	(None, 7972, 16)	0	batch_normalization_1[0][0]
conv1d_2 (Conv1D)	(None, 7972, 16)	784	activation_1[0][0]
batch_normalization_2 (Batch_normalization)	(None, 7972, 16)	64	conv1d_2[0][0]
activation_2 (Activation)	(None, 7972, 16)	0	batch_normalization_2[0][0]
conv1d_3 (Conv1D)	(None, 7972, 16)	784	activation_2[0][0]
batch_normalization_3 (Batch_normalization)	(None, 7972, 16)	64	conv1d_3[0][0]
activation_3 (Activation)	(None, 7972, 16)	0	batch_normalization_3[0][0]
conv1d_4 (Conv1D)	(None, 7972, 16)	784	activation_3[0][0]
batch_normalization_4 (Batch_normalization)	(None, 7972, 16)	64	conv1d_4[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

add_1 (Add)	(None, 7972, 16)	0	batch_normalization_4[0][0] batch_normalization_2[0][0]
activation_4 (Activation)	(None, 7972, 16)	0	add_1[0][0]
conv1d_5 (Conv1D)	(None, 7972, 16)	784	activation_4[0][0]
batch_normalization_5 (Batch_normalization)	(None, 7972, 16)	64	conv1d_5[0][0]
activation_5 (Activation)	(None, 7972, 16)	0	batch_normalization_5[0][0]
conv1d_6 (Conv1D)	(None, 7972, 16)	784	activation_5[0][0]
batch_normalization_6 (Batch_normalization)	(None, 7972, 16)	64	conv1d_6[0][0]
activation_6 (Activation)	(None, 7972, 16)	0	batch_normalization_6[0][0]
conv1d_7 (Conv1D)	(None, 7972, 16)	784	activation_6[0][0]
batch_normalization_7 (Batch_normalization)	(None, 7972, 16)	64	conv1d_7[0][0]
activation_7 (Activation)	(None, 7972, 16)	0	batch_normalization_7[0][0]
conv1d_8 (Conv1D)	(None, 7972, 16)	784	activation_7[0][0]
batch_normalization_8 (Batch_normalization)	(None, 7972, 16)	64	conv1d_8[0][0]
add_2 (Add)	(None, 7972, 16)	0	batch_normalization_8[0][0] batch_normalization_5[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_8 (Activation)	(None, 7972, 16)	0	add_2[0][0]
conv1d_9 (Conv1D)	(None, 3986, 32)	1568	activation_8[0][0]
batch_normalization_9 (Batch_normalization)	(None, 3986, 32)	128	conv1d_9[0][0]
activation_9 (Activation)	(None, 3986, 32)	0	batch_normalization_9[0][0]
conv1d_10 (Conv1D)	(None, 3986, 32)	3104	activation_9[0][0]
conv1d_11 (Conv1D)	(None, 3986, 32)	3104	conv1d_10[0][0]
batch_normalization_10 (Batch_normalization)	(None, 3986, 32)	128	conv1d_11[0][0]
add_3 (Add)	(None, 3986, 32)	0	batch_normalization_10[0][0] batch_normalization_9[0][0]
activation_10 (Activation)	(None, 3986, 32)	0	add_3[0][0]
conv1d_12 (Conv1D)	(None, 3986, 32)	3104	activation_10[0][0]
batch_normalization_11 (Batch_normalization)	(None, 3986, 32)	128	conv1d_12[0][0]
activation_11 (Activation)	(None, 3986, 32)	0	batch_normalization_11[0][0]
conv1d_13 (Conv1D)	(None, 3986, 32)	3104	activation_11[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_12 (Batch_normalization)	(None, 3986, 32)	128	conv1d_13[0][0]
add_4 (Add)	(None, 3986, 32)	0	batch_normalization_12[0][0] batch_normalization_11[0][0]
activation_12 (Activation)	(None, 3986, 32)	0	add_4[0][0]
conv1d_14 (Conv1D)	(None, 3986, 32)	3104	activation_12[0][0]
batch_normalization_13 (Batch_normalization)	(None, 3986, 32)	128	conv1d_14[0][0]
activation_13 (Activation)	(None, 3986, 32)	0	batch_normalization_13[0][0]
conv1d_15 (Conv1D)	(None, 3986, 32)	3104	activation_13[0][0]
batch_normalization_14 (Batch_normalization)	(None, 3986, 32)	128	conv1d_15[0][0]
add_5 (Add)	(None, 3986, 32)	0	batch_normalization_14[0][0] batch_normalization_13[0][0]
activation_14 (Activation)	(None, 3986, 32)	0	add_5[0][0]
conv1d_16 (Conv1D)	(None, 1993, 64)	6208	activation_14[0][0]
batch_normalization_15 (Batch_normalization)	(None, 1993, 64)	256	conv1d_16[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

activation_15 (Activation)	(None, 1993, 64)	0	batch_normalization_15[0][0]
conv1d_17 (Conv1D)	(None, 1993, 64)	12352	activation_15[0][0]
conv1d_18 (Conv1D)	(None, 1993, 64)	12352	conv1d_17[0][0]
batch_normalization_16 (Batch_normalization)	(None, 1993, 64)	256	conv1d_18[0][0]
add_6 (Add)	(None, 1993, 64)	0	batch_normalization_16[0][0] batch_normalization_15[0][0]
activation_16 (Activation)	(None, 1993, 64)	0	add_6[0][0]
conv1d_19 (Conv1D)	(None, 1993, 64)	12352	activation_16[0][0]
batch_normalization_17 (Batch_normalization)	(None, 1993, 64)	256	conv1d_19[0][0]
activation_17 (Activation)	(None, 1993, 64)	0	batch_normalization_17[0][0]
conv1d_20 (Conv1D)	(None, 1993, 64)	12352	activation_17[0][0]
batch_normalization_18 (Batch_normalization)	(None, 1993, 64)	256	conv1d_20[0][0]
add_7 (Add)	(None, 1993, 64)	0	batch_normalization_18[0][0] batch_normalization_17[0][0]
activation_18 (Activation)	(None, 1993, 64)	0	add_7[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

conv1d_21 (Conv1D)	(None, 1993, 64)	12352	activation_18[0][0]
batch_normalization_19 (Batch_normalization)	(None, 1993, 64)	256	conv1d_21[0][0]
activation_19 (Activation)	(None, 1993, 64)	0	batch_normalization_19[0][0]
conv1d_22 (Conv1D)	(None, 1993, 64)	12352	activation_19[0][0]
batch_normalization_20 (Batch_normalization)	(None, 1993, 64)	256	conv1d_22[0][0]
add_8 (Add)	(None, 1993, 64)	0	batch_normalization_20[0][0] batch_normalization_19[0][0]
activation_20 (Activation)	(None, 1993, 64)	0	add_8[0][0]
flatten_1 (Flatten)	(None, 127552)	0	activation_20[0][0]
dense1 (Dense)	(None, 2048)	261228544	flatten_1[0][0]
batch_normalization_21 (Batch_normalization)	(None, 2048)	8192	dense1[0][0]
dropout1 (Dropout)	(None, 2048)	0	batch_normalization_21[0][0]
activation_21 (Activation)	(None, 2048)	0	dropout1[0][0]
dense5 (Dense)	(None, 1024)	2098176	activation_21[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_22 (Batch_normalization)	(None, 1024)	4096	dense5[0][0]
dropout5 (Dropout)	(None, 1024)	0	batch_normalization_22[0][0]
activation_22 (Activation)	(None, 1024)	0	dropout5[0][0]
dense6 (Dense)	(None, 512)	524800	activation_22[0][0]
batch_normalization_23 (Batch_normalization)	(None, 512)	2048	dense6[0][0]
dropout6 (Dropout)	(None, 512)	0	batch_normalization_23[0][0]
activation_23 (Activation)	(None, 512)	0	dropout6[0][0]
dense7 (Dense)	(None, 1024)	525312	activation_23[0][0]
batch_normalization_24 (Batch_normalization)	(None, 1024)	4096	dense7[0][0]
dropout7 (Dropout)	(None, 1024)	0	batch_normalization_24[0][0]
add_9 (Add)	(None, 1024)	0	dropout7[0][0] batch_normalization_22[0][0]
activation_24 (Activation)	(None, 1024)	0	add_9[0][0]
dense8 (Dense)	(None, 512)	524800	activation_24[0][0]

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

batch_normalization_25 (Batch_normalization)	(None, 512)	2048	dense8[0][0]
dropout8 (Dropout)	(None, 512)	0	batch_normalization_25[0][0]
activation_25 (Activation)	(None, 512)	0	dropout8[0][0]
dense9 (Dense)	(None, 256)	131328	activation_25[0][0]
batch_normalization_26 (Batch_normalization)	(None, 256)	1024	dense9[0][0]
dropout9 (Dropout)	(None, 256)	0	batch_normalization_26[0][0]
activation_26 (Activation)	(None, 256)	0	dropout9[0][0]
dense10 (Dense)	(None, 128)	32896	activation_26[0][0]
batch_normalization_27 (Batch_normalization)	(None, 128)	512	dense10[0][0]
dropout10 (Dropout)	(None, 128)	0	batch_normalization_27[0][0]
activation_27 (Activation)	(None, 128)	0	dropout10[0][0]
predictions (Dense)	(None, 1)	129	activation_27[0][0]
Total params: 265,196,881 **Trainable params: 265,184,465**			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Non-trainable params: 12,416			

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S17. Parameters of ResNet models in all scenarios.

Scenarios	Total parameters	Total number of layers	The number of convolution layers	The number of fully connected layers	The number of skip connections	Architecture	Loss function	Optimizer	Learning rate	Training epoch	Batch size
Scenario 1 (MDG-160K)	81,171,793	30	22	8	9	Supplementary Table S10 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 2 (EDC-11K)	73,831,761	30	22	8	9	Supplementary Table S11 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 3 (EYDC-9K)	140,023,121	30	22	8	9	Supplementary Table S12 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 4 (MDC-9K)	81,171,793	30	22	8	9	Supplementary Table S10 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 5 (ODC-11K)	67,409,233	30	22	8	9	Supplementary Table S13 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 6 (MEDG-61K)	143,299,921	30	22	8	9	Supplementary Table S14 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 7 (MEYDG-51K)	209,491,281	30	22	8	9	Supplementary Table S15 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 8 (MEYODG-51K)	265,196,881	30	22	8	9	Supplementary Table S16 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 9 (MEDC-7K)	143,299,921	30	22	8	9	Supplementary Table S14 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 10 (MEYDC-6K)	209,491,281	30	22	8	9	Supplementary Table S15 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100
Scenario 11 (MEYODC-6K)	265,196,881	30	22	8	9	Supplementary Table S16 and Supplementary Figure S3	Mean square error (MSE)	Adam	0.0002	150	100

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S18. ML models and their parameters in all scenarios.

Models	Scenario 1* (MDG-160K)	Scenario 2 (EDC-11K)	Scenario 3 (EYDC-9K)	Scenario 4 (MDC-9K)	Scenario 5 (ODC-11K)	Scenario 6 (MEDG-61K)	Scenario 7 (MEYDG-51K)	Scenario 8 (MEYODG-51K)	Scenario 9 (MEDC-7K)	Scenario 10 (MEYDC-6K)	Scenario 11 (MEYODC-6K)
Lasso	NA	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001
Ridge	NA	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001	Alpha: 0.001
SVR	C: 0.1	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01	C: 0.01
XGBoost	The number of trees: 200	Default	Default	Default	Default	Default	Default	Default	Default	Default	Default
Random forest	The number of trees: 150 The depth of tree: 8	Default	Default	Default	Default	Default	Default	Default	Default	Default	Default

NA: not available

* Hyperparameter optimization in ML was conducted in scenario 1

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S19. Prediction performance of SVR in scenario 1 (MDG-160K) in a set of hyperparameter values.

SVR		
C parameter	**RMSE**	**R^2**
0.01	1.304	0.782
0.1	1.304	0.782
1 (Default)	1.305	0.78
5	1.69	0.634
10	2.107	0.431

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S20. Prediction performance of random forest in scenario 1 (MDG-160K) in a set of hyperparameter values.

Random forest		
The number of trees	**RMSE**	**R^2**
50	1.429	0.738
100 (Default)	1.602	0.671
150	1.427	0.739
200	1.427	0.739
250	2.108	0.43

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S21. Prediction performance of XGBoost in scenario 1 (MDG-160K) in a set of hyperparameter values.

XGBoost			
The number of trees	**the depth of tree**	**RMSE**	**R^2**
50	**6 (default)**	1.43	0.738
100 (Default)	**6 (default)**	1.486	0.716
150	**6 (default)**	1.354	0.765
50	**4**	1.486	0.717
100 (Default)	**4**	1.43	0.738
150	**4**	1.397	0.75
50	**8**	1.406	0.746
100 (Default)	**8**	1.356	0.764
150	**8**	1.328	0.774

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S22. Prediction performance of ML models in scenario 1 (MDG-160K) with feature selection.

Scenario 1 Feature selection	% of selected feature	RMSE	R^2
Random forest	30%	1.415	0.743
	50%	1.419	0.742
	70%	1.435	0.736
	100%	1.427	0.739
SVR	30%	1.307	0.781
	50%	1.304	0.782
	70%	1.304	0.782
	100%	1.304	0.782
XGBoost	30%	1.333	0.772
	50%	1.329	0.773
	70%	1.328	0.774
	100%	1.328	0.774

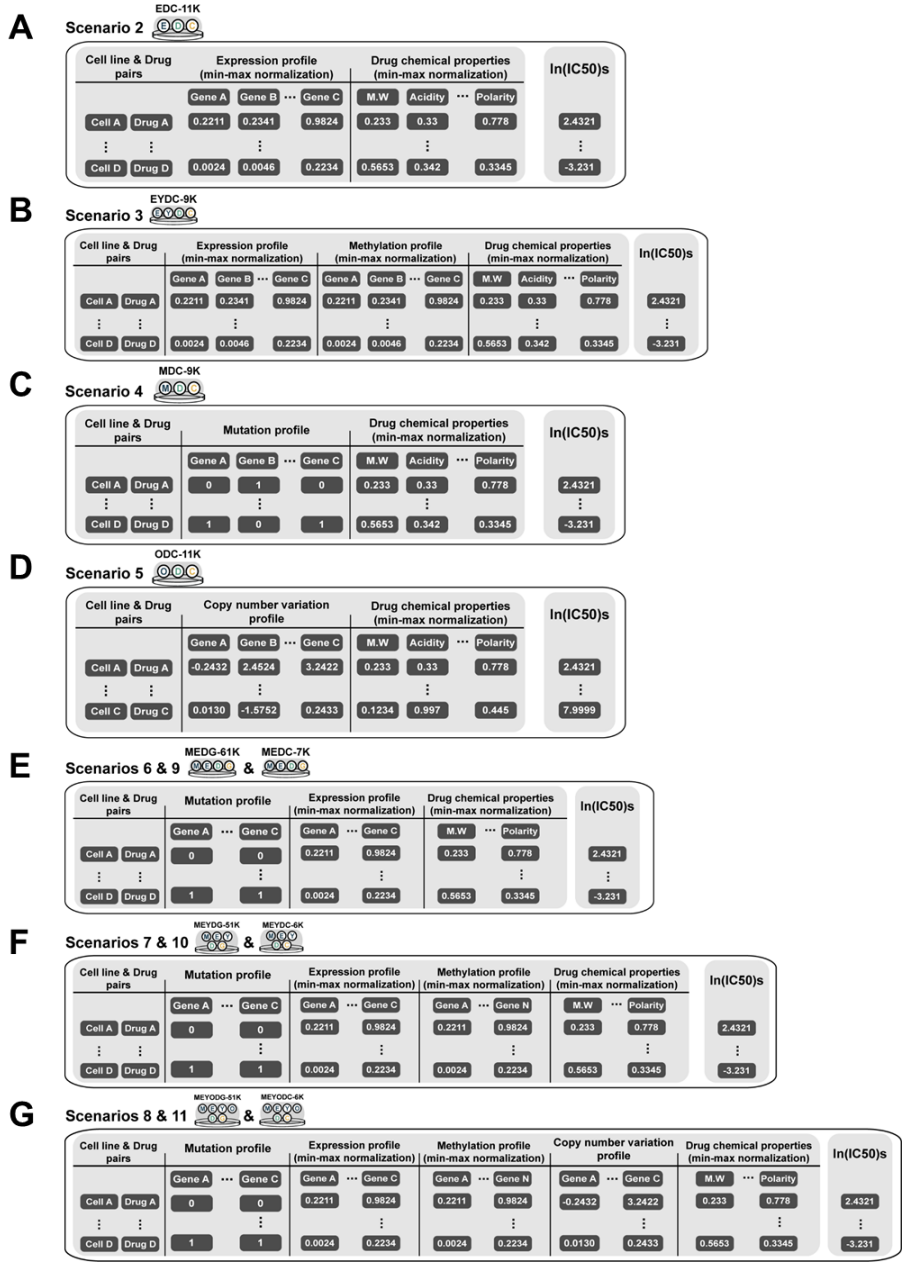
Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Table S23. Prediction performance of ML models with feature selection in scenario 4 (MDC-9K).

Scenario 4 Feature selection	% of selected feature	RMSE	R^2
Lasso	30%	1.041	0.702
	70%	1.04	0.702
	100%	1.04	0.702
Ridge	30%	1.052	0.695
	70%	1.054	0.694
	100%	1.051	0.696
SVR	30%	1.546	0.342
	70%	1.644	0.256
	100%	1.646	0.254
Random forest	30%	1.065	0.688
	70%	1.153	0.634
	100%	1.152	0.635
XGBoost	30%	1.032	0.707
	70%	1.032	0.707
	100%	1.032	0.707

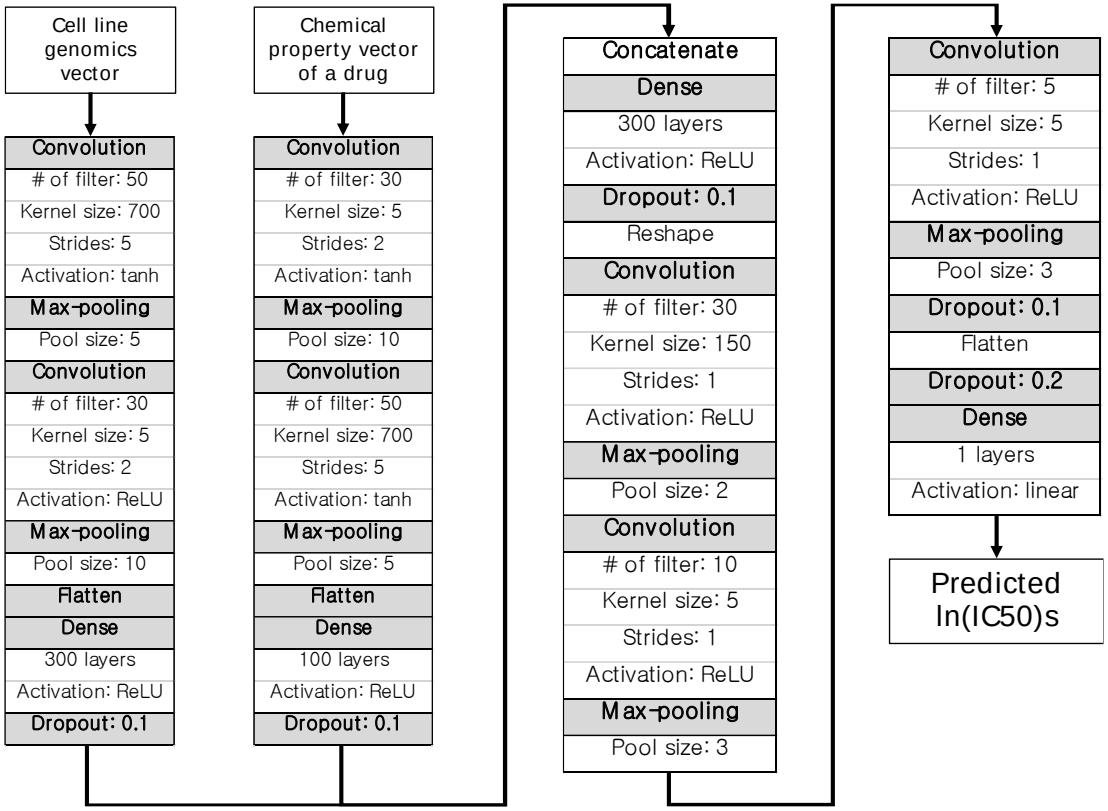
Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Figure S1. Input vector description for scenarios 2 to 11 in unified models. Input vector contains various combinations of genomic profiles (mutation, expression, DNA methylation, and copy number variation), and drug chemical properties. (a) scenario 2 (EDC-11K) (b) scenario 3 (EYDC-9K). (c) scenario 4 (MDC-9K). (d) scenario 5 (ODC-11K). (e) scenarios 6 and 9 (MEDG-61K and MEDC-7K) (f) scenarios 7 and 10 (MEYDG-51K and MEYDC-6K). (g) scenario 8 and 11 (MEYODG-51K and MEYODC-6K). Column ln(IC50)s represents output variable.



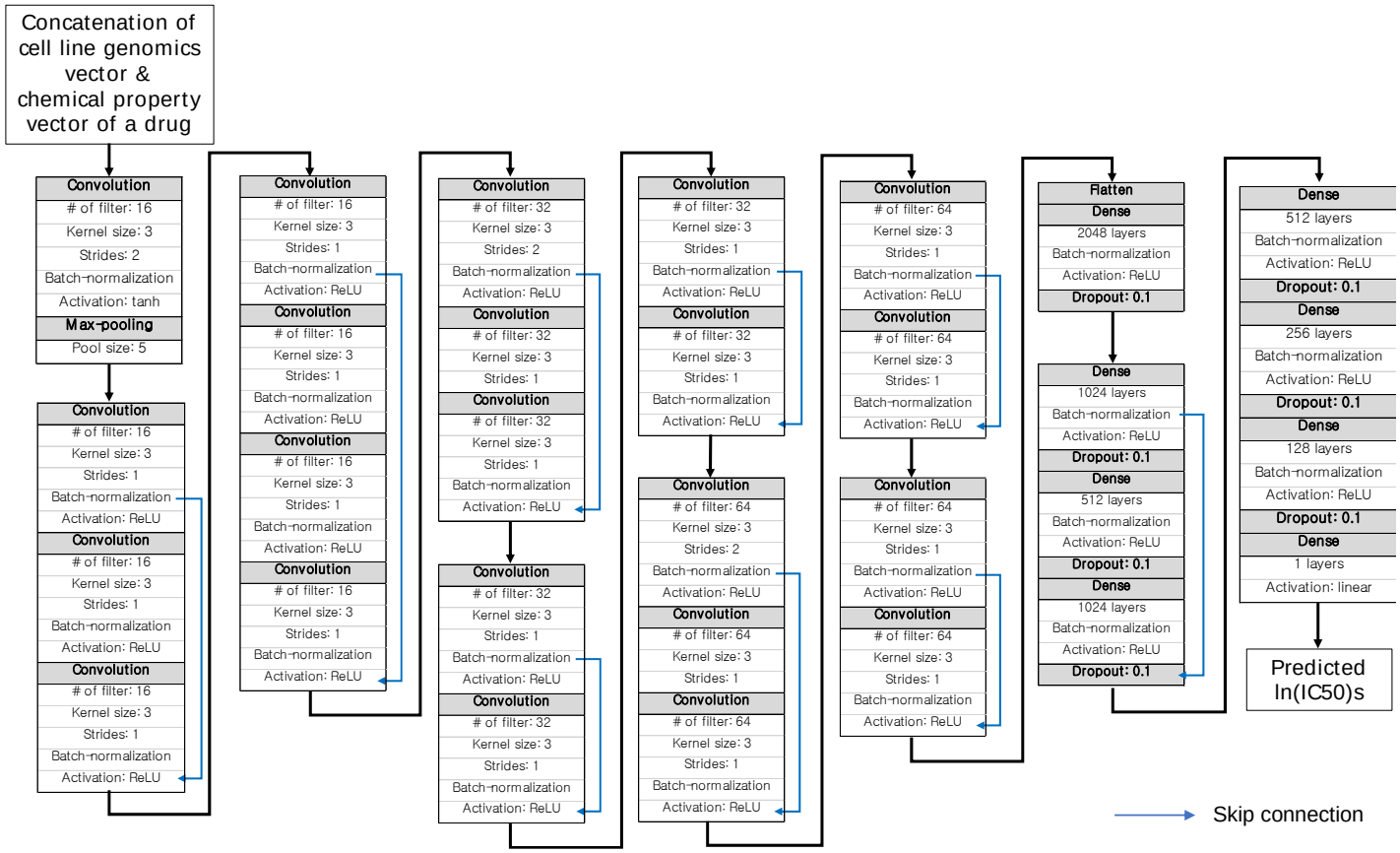
Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Figure S2. The architecture of the original CDRScan. Grey boxes represent convolution and fully-connected layers. Cell line genomics vector (e.g., mutation, gene expression, DNA methylation, and copy number variation) and a drug chemical property vector were assigned as the two individual input vectors in the first layer. The original CDRScan only considers a mutation status vector of a cell line in genomics data.



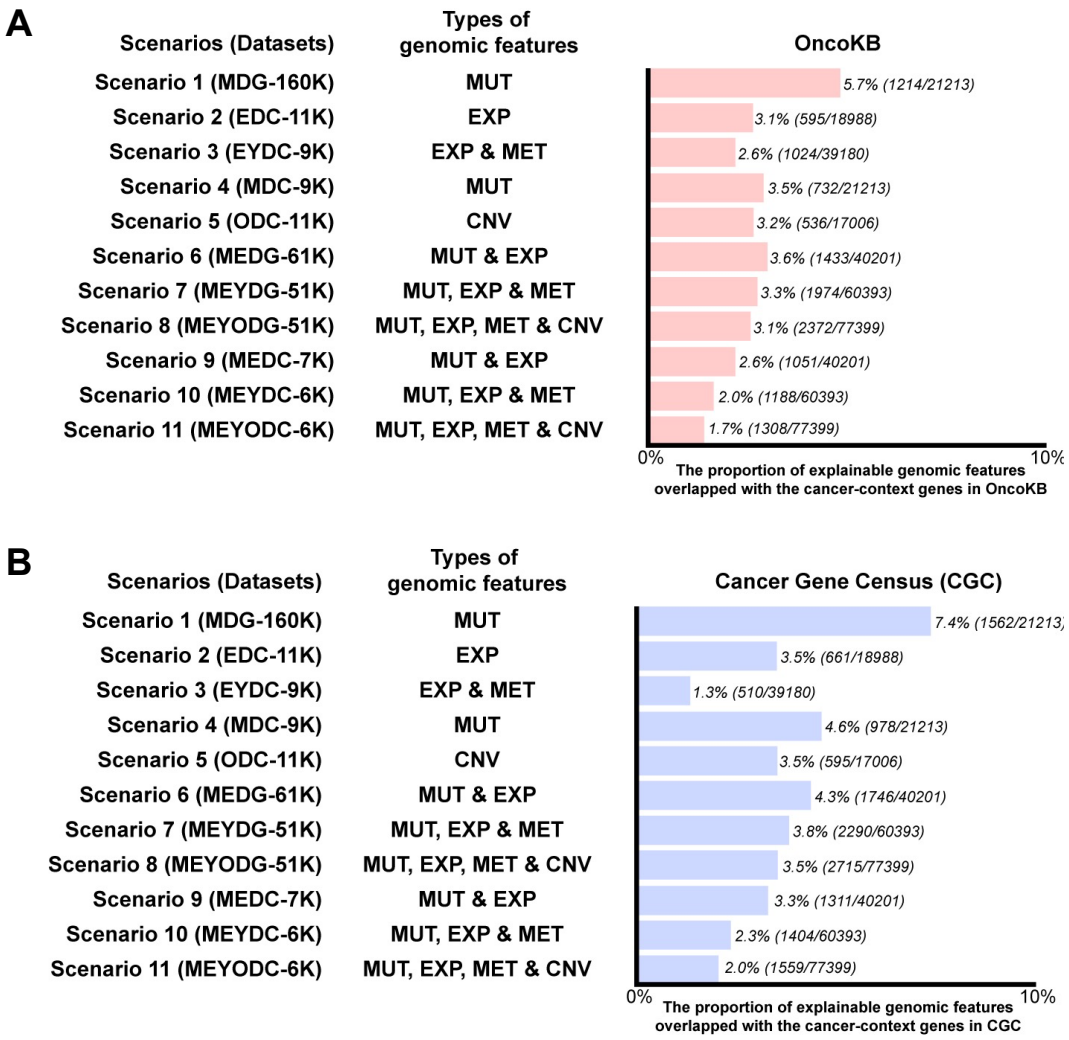
Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Figure S3. ResNet structure including skip connections (indicated in blue arrow). Grey boxes represent convolution and fully-connected layers. Concatenation of cell line genomics vector (e.g., mutation, gene expression, DNA methylation, and copy number variation) and drug chemical property vector was used as input layer.



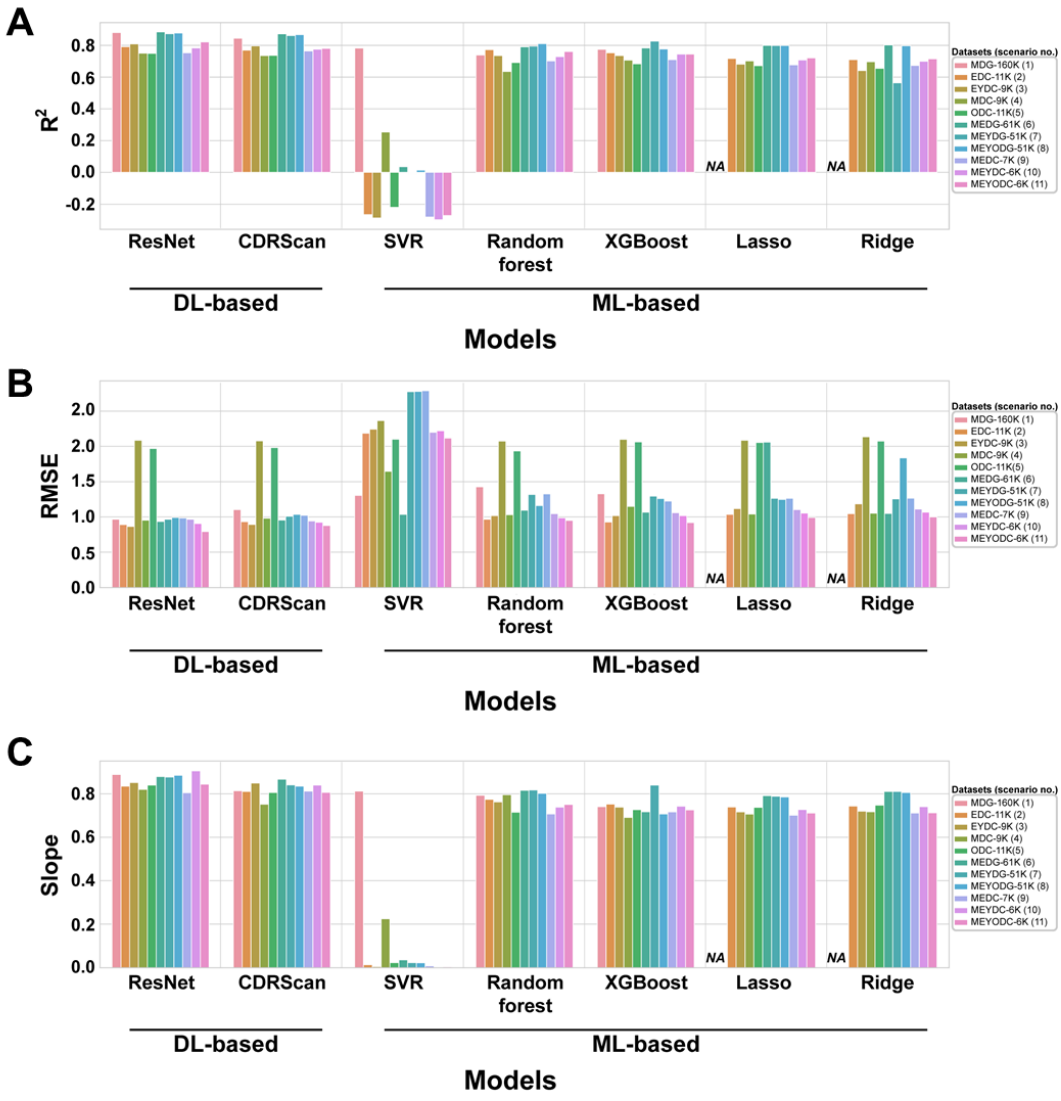
Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Figure S4. The proportion of explainable genomic features overlapped with cancer-context genes over the total number of genomic features. We obtained explainable genomic features by applying LIME to ResNet models in eleven scenarios. We inspected how many explainable genomic features overlapped with the cancer-context genes in each ResNet model. Subsequently, we calculated the proportion by dividing the number of explainable genomic features overlapped with cancer-context genes by the total number of genomic features. The proportion of explainable genomic features overlapped with the cancer-context genes from (a) the OncoKB (Chakravarty, et al., 2017) and (b) the Cancer Gene Census (CGC) (Sondka, et al., 2018) were represented, respectively. MUT (mutations), EXP (expression), MET (DNA methylation), and CNV (copy number variations).



Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary Figure S5. Summary of performance comparisons of the models, in the test sets of all the scenarios, in terms of (a) R^2, (b) RMSE, and (c) slope. The y-axis represents R^2, RMSE and slope values, and the x-axis the models. It is noted that lasso and ridge were not available in scenario 1, due to exhaustive computation times. The ResNet trained in scenario 1 is called as ResNetIC50 throughout the manuscript. NA: not available.



Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024

Supplementary References

Cancer Genome Atlas Research Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature, 490*(7418), 61-70. doi:10.1038/nature11412

Chakravarty, D.*, et al.* (2017). OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.*(1), 1-16. doi:10.1200/PO.17.00011

Hastie T, T. R., Narasimhan B, Chu G. (2019). impute: impute: Imputation for microarray data. R package version 1.58.0.

Lind, A. P., & Anderson, P. C. (2019). Predicting drug activity against cancer cells by random forest models based on minimal genomic information and chemical properties. *PLoS One, 14*(7), e0219774. doi:10.1371/journal.pone.0219774

Sondka, Z.*, et al.* (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer., 18*(11), 696-705. doi:10.1038/s41568-018-0060-1

Yap, C. W. (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J. Comput. Chem., 32*(7), 1466-1474. doi:10.1002/jcc.21707

Downloaded from https://academic.oup.com/bioinformatics/article/38/10/2810/6553012 by guest on 10 May 2024