Supplementary

BioGCN: Biologically Inspired Graph Convolutional Network to Predict Human Oral Bioavailability

Koyel Mandal[1,2], Aritra Banerjee[1], Anushree Bhattacharya[1], and Sanghamitra Bandyopadhyay[1,*]

[1]Indian Statistical Institute, Machine Intelligence Unit, Kolkata, 700108, West Bengal, India
[1]Departmnet of Computer Science and Engineering, SRM University-AP, Amaravati 522 240, Andhra Pradesh, India
[*]sanghami@isical.ac.in

Table S1: Detailed description of test dataset 1.

Drug Name	Bioavailability	Class
Baclofen	70	1
Norfenfluramine	85	1
Stavudine	82	1
Chlorpropamide	95	1
Sulthiame	73	1
Ganciclovir	4	0
Trapidil	96	1
Rilmenidine	100	1
Cimetidine	60	1
Guanadrel	85	1
Disulfiram	80	1
Guanethidine	20	1
Flosequinan	72	1
Flutamide	90	1
Piroximone	81	1
Sulfapyridine	60	1
Chlormezanone	99	1
Sulfamerazine	81	1
Butalbital	60	1
Amobarbital	95	1
Neramexane	90	1
Thiocyanate	99	1
Azosemide	10	0
Primidone	92	1
Sulfamethazine	86.5	1
Procarbazine	100	1
Memantine	99	1
Leflunomide	80	1
Milrinone	86	1
Ethacrynic	49	1
Nimesulide	70	1

Drug Name	Bioavailability	Class
Brodimoprim	80	1
Dexmedetomidine	20	1
Oxantel	0	0
Melatonin	12	0
Aminopyrine	100	1
Metamizol	85	1
Procaine	0	0
Toliprolol	90	1
Timolol	61	1
Dantrolene	70	1
Meclofenamic	60	1
Suprofen	90	1
Naproxen	99	1
Brofaromine	90	1
Frovatriptan	27	1
Physostigmine	10	0
Methapyrilene	20	1
Methylphenidate	22	1
Clopamide	85	1
Rivastigmine	72	1
Valganciclovir	61	1
Oxazepam	97	1
Bromfenac	67	1
Carbamazepine	70	1
Idronoxil	1	0
Oxcarbazepine	95	1
Bendroflumethiazide	90	1
Mefenamic	49	1
Flupirtine	90	1
Enoxacin	87	1
Cefmetazole	80	1
Raclopride	65	1
Gliclazide	97	1
Mepivacaine	55	1
Artemisinin	51	1
Ciramadol	82	1
Doripenem	0	0
Fluindione	81	1
Clorazepate	91	1
Flunitrazepam	80	1
Mazindol	93	1
Temazepam	75	1
Voriconazole	96	1
Ketoprofen	90	1
Azidocillin	57	1
Sulbenicillin	0	0
Piribedil	10	0
Cefroxadine	100	1
Medifoxamine	21	1
Pheniramine	99	1
Remoxipride	90	1
Carteolol	88	1
Xamoterol	7	0
Tramadol	72.5	1
Cilastatin	0	0

Drug Name	Bioavailability	Class
Ponalrestat	90	1
Alprazolam	88	1
Roflumilast	62	1
Etizolam	87	1
Cefixime	51	1
Indoprofen	100	1
Clinafloxacin	90	1
Ondansetron	62	1
Fluoxetine	60	1
Ciprofloxacin	60	1
Rufloxacin	70	1
Piretanide	92	1
Alosetron	55	1
Omeprazole	53	1
Morphine	24	1
Meticillin	0	0
Diphenhydramine	72	1
Phencyclidine	72	1
Almotriptan	70	1
Levobunolol	75	1
Trospectomycin	7	0
Picosulfate	10	0
Loxapine	30	1
Chlorprothixene	41	1
Felodipine	15	0
Mianserin	22	1
Nimodipine	11	0
Hydrocodone	80	1
Phenoxybenzamine	25	1
Amosulalol	99	1
Granisetron	60	1
Prucalopride	90	1
Glibornuride	91	1
Bupivacaine	49	1
Bambuterol	15	0
Betaxolol	85	1
Pafenolol	27	1
Dicloxacillin	67.5	1
Pimobendan	60	1
Nilvadipine	14	0
Paroxetine	50	1
Aminopterin	84	1
Oxmetidine	53	1
Prazosin	68	1
Sparfloxacin	92	1
Grepafloxacin	70	1
Oxyfedrine	85	1
Reboxetine	94	1
Terazosin	90	1
Formestane	20	1
Cicloprolol	100	1
Elinogrel	50	1
Flucloxacillin	60	1
Fenofibrate	60	1
Papaverine	51	1

Drug Name	Bioavailability	Class
Repirinast	5	0
Tianeptine	99	1
Epanolol	8	0
Azlocillin	0	0
Quinine	75	1
Hycanthone	0	0
Lisuride	14	0
Terodiline	82	1
Cilostazol	87	1
Glisoxepid	95	1
Diacetolol	35	1
Methyltestosterone	50	1
Deramciclane	23	1
Kanamycin	1	0
Escitalopram	80	1
Raltitrexed	15	0
Panobinostat	30	1
Heroin	1	0
Mezlocillin	0	0
Thioridazine	40	1
Delavirdine	85	1
Gestodene	93	1
Hydroxyzine	80	1
Chlormadinone	5	0
Nicainoprol	70	1
Ifenprodil	5	0
Cimetropium	2	0
Butorphanol	17	0
Loxiglumide	97	1
Butylscopolamine	1	0
Efegatran	6	0
Mesterolone	3	0
Atovaquone	23	1
Droperidol	75	1
Ketanserin	50	1
Chlortetracycline	30	1
Losartan	35	1
Noscapine	30	1
Gefitinib	60	1
Oxytetracycline	58	1
Domperidone	15	0
Lobeline	0	0
Prifinium_bromide	3	0
Dexamethasone	78	1
Pivampicillin	31	1
Chlorhexidine	1	0
Pirmenol	83	1
Avitriptan	17.200001	0
Sildenafil	38	1
Desogestrel	5	0
Methylprednisolone	82	1
Cisapride	40	1
Opipramol	95	1
Eproxindine	70	1
Vardenafil	15	0

Drug Name	Bioavailability	Class
Estradiol_17-valerate	3	0
Finasteride	63	1
Argatroban	0	0
Abecarnil	65	1
Nicergoline	5	0
Sertindole	74	1
Loteprednol	0	0
Abiraterone	0	0
Glimepiride	100	1
Nexopamil	23	1
Dipyridamole	50	1
Pitavastatin	51	1
Dabigatran	6.5	0
Irbesartan	70	1
Picumast	57	1
Nufenoxole	100	1
Metergoline	23	1
Quinapril	49	1
Fluticasone_propionate	1	0
Deflazacort	68	1
Silodosin	32	1
Budesonide	37	1
Amprenavir	35	1
Epristeride	93	1
Simvastatin	3	0
Acarbose	2	0
Carindacillin	69	1
Lofepramine	10	0
Idarubicin	28	1
Lacidipine	18	0
Clofazimine	57	1
Pranlukast	20	1
Nandrolone	3	0
Verapamil	22	1
Calcipotriol	5.5	0
Lurtotecan	6	0
Hesperetin	20	1
Lurasidone	19	0
Tiropramide	23	1
Ergocalciferol	55	1
Lapatinib	51	1
Pyronaridine	20	1
Tigecycline	0	0
Maraviroc	23	1
Ouabain	1	0
Acetohydroxamic acid	55	1
Bazedoxifene	6	0
Ulipristal	56	1
Dihydroergosine	10	0
Fulvestrant	0	0
Teniposide	41	1
Ergotamine	2	0
Irinotecan	8	0
Bromocriptine	28	1
Montelukast	62	1

Drug Name	Bioavailability	Class
Indinavir	65	1
Erythromycin	35	1
Saquinavir	4	0
Spiramycin	35	1
Desmopressin	0.1	0
Nystatin	1	0
Fomepizole	100	1
Mivacurium	0	0
Methionine	100	1
Amifostine	99	1
Aminolevulinic_acid	60	1
Porfimer	0	0
Trientine	92.5	1
Isoniazid	80	1
Gaboxadol	84	1
Propylthiouracil	85	1
Chlorzoxazone	100	1
Hydrochlorothiazide	71	1
pralidoxime	51	1
Lazabemide	78	1
Norepinephrine	3	0
Cyclic-hpmpc	0	0
Tinidazole	98	1
Lipoic	29	1
Tranexamic_acid	54	1
Riluzole	64	1
Trichlormethiazide	60	1
Hydroflumethiazide	50	1
Hydralazine	23	1
Acefylline	25	1
Uridine	8	0
Etofylline	80	1
Perzinfotel	35	1
Zalcitabine	88	1
Chloroxine	89	1
Foscarnet	9	0
Tirapazamine	75	1

Table S2: Detailed description of test dataset 2.

Drug Name	Bioavailability	Class
Vardenafil	15	0
Sildenafil	25	1
Propafenone	90	1
Venlafaxine	45	1
Naltrexone	20	1
Lurasidone	10	0
Plazomicin	0	0
Vonoprazan fumarate	70	1
Trifarotene	18	0
Dalbavancin	7	0
Icaridin	0	0
Triamcinolone acetonide	23	1
Sugammadex	0	0

Drug Name	Bioavailability	Class
Gadoteridol acetate	0	0
Biapenem	2	0
Pixantrone	0	0
Denaverine	37	1
Pazopanib	21	1
Fipronil	0.9	0
Olaparib	25	1
Dronedarone	15	0
Naproxen sodium	95	1
Afatinib	60	1
Atracurium	100	1
Besifloxacin	80	1
Citicholine	90	1
Cinitapride	100	1
Atazanavir	60	1
Cefepime	82	1
Sorafenib	38	1
Ozagrel hcl	100	1
Daptomycin	0	0
Tiotropium	2	0
Colistin	4	0
Methoxsalen	50	1
Trometamol	40	1
Lenalidomide	82	1
Thiocolchicoside	25	1
Doxercalciferol	42	1
Loxoprofen sodium	73	1
Risedronate sodium	0.63	0
Tolperisone	17	0
Cabergoline	30	1
Testolactone	100	1
Fenfluramine	68	1

Molecular descriptors: The total number of descriptors used in this work is 200 using the RDKit Python package. The detailed description of the molecular descriptors is given in Table S3. After the preprocessing step as mentioned in the main paper, the reduced number of molecular descriptors is 44. They are BalabanJ, BertzCT, Chi0, EState_VSA10, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA8, FpDensityMorgan1, FractionCSP3, HallKierAlpha, Ipc, Kappa3, MaxAbsEStateIndex, MaxAbsPartialCharge, MaxPartialCharge, MinAbsEStateIndex, MinEStateIndex, MinPartialCharge, MolLogP, NHOHCount, NOCount, NumAromaticRings, PEOE_VSA1, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, RingCount, SMR_VSA1, SMR_VSA10, SMR_VSA5, SMR_VSA6, SMR_VSA7, SlogP_VSA3, SlogP_VSA5, VSA_EState2, VSA_EState3, VSA_EState4, VSA_EState5, VSA_EState6, VSA_EState7, VSA_EState8, and qed.

Table S3: The detailed description of molecular descriptors used in RDKit Package.

Type of Descriptors	No. of Descriptors	Descriptors
Topological	18	BalabanJ, BertzCT, Chi0, Chi0n, Chi0v, Chi1, Chi1n, Chi1v, Chi2n, Chi2v, Chi3n, Chi3v, Chi4n, Chi4v, HallKierAlpha, Kappa1, Kappa2, Kappa3

Type of Descriptors	No. of Descriptors	Descriptors
Basic EState	4	MaxAbsEStateIndex, MaxEStateIndex, MinEStateIndex, MinAbsEStateIndex
Hybrid EState-VSA	21	EState_VSA1, EState_VSA2, EState_VSA3, EState_VSA4, EState_VSA5, EState_VSA6, EState_VSA7, EState_VSA8, EState_VSA9, EState_VSA10, EState_VSA11, VSA_EState1, VSA_EState2, VSA_EState3, VSA_EState4, VSA_EState5, VSA_EState6, VSA_EState7, VSA_EState8, VSA_EState9, VSA_EState10
Quantitative estimation of drug-likeness	1	qed
General Physicochemical	11	MinAbsPartialCharge, NumRadicalElectrons, FpDensityMorgan1, FpDensityMorgan2, FpDensityMorgan3, HeavyAtomMolWt, MaxAbsPartialCharge, MinPartialCharge, ExactMolWt, NumValenceElectrons, MaxPartialCharge
Crippen	2	MolLogP, MolMR
Molecular surface area	38	LabuteASA, PEOE_VSA1, PEOE_VSA2, PEOE_VSA3, PEOE_VSA4, PEOE_VSA5, PEOE_VSA6, PEOE_VSA7, PEOE_VSA8, PEOE_VSA9, PEOE_VSA10, PEOE_VSA11, PEOE_VSA12, PEOE_VSA13, PEOE_VSA14, SMR_VSA1, SMR_VSA2, SMR_VSA3, SMR_VSA4, SMR_VSA5, SMR_VSA6, SMR_VSA7, SMR_VSA8, SMR_VSA9, SMR_VSA10, SlogP_VSA1, SlogP_VSA2, SlogP_VSA3, SlogP_VSA4, SlogP_VSA5, SlogP_VSA6, SlogP_VSA7, SlogP_VSA8, SlogP_VSA9, SlogP_VSA10, SlogP_VSA11, SlogP_VSA12, TPSA
Fragments	85	fr_Al_COO, fr_Al_OH, fr_Al_OH_noTert, fr_ArN, fr_Ar_COO, fr_Ar_N, fr_Ar_NH, fr_Ar_OH, fr_COO, fr_COO2, fr_C_O, fr_C_O_noCOO, fr_C_S, fr_HOCCN, fr_Imine, fr_NH0, fr_NH1, fr_NH2, fr_N_O, fr_Ndealkylation1, fr_Ndealkylation2, fr_Nhpyrrole, fr_SH, fr_aldehyde, fr_alkyl_carbamate, fr_alkyl_halide, fr_allylic_oxid, fr_amide, fr_amidine, fr_aniline, fr_aryl_methyl, fr_azide, fr_azo, fr_barbitur, fr_benzene, fr_benzodiazepine, fr_bicyclic, fr_diazo, fr_dihydropyridine, fr_epoxide, fr_ester, fr_ether, fr_furan, fr_guanido, fr_halogen, fr_hdrzine, fr_hdrzone, fr_imidazole, fr_imide, fr_isocyan, fr_isothiocyan, fr_ketone, fr_ketone_Topliss, fr_lactam, fr_lactone, fr_methoxy, fr_morpholine, fr_nitrile, fr_nitro, fr_nitro_arom, fr_nitro_arom_nonortho, fr_nitroso, fr_oxazole, fr_oxime, fr_para_hydroxylation, fr_phenol, fr_phenol_noOrthoHbond, fr_phos_acid, fr_phos_ester, fr_piperdine, fr_piperzine, fr_priamide, fr_prisulfonamd, fr_pyridine, fr_quatN, fr_sulfide, fr_sulfonamd, fr_sulfone, fr_term_acetylene, fr_tetrazole, fr_thiazole, fr_thiocyan, fr_thiophene, fr_unbrch_alkane, fr_urea
Lipinski	18	FractionCSP3, HeavyAtomCount, NHOHCount, NOCount, NumAliphaticCarbocycles, NumAliphaticHeterocycles, NumAliphaticRings, NumAromaticCarbocycles, NumAromaticHeterocycles, NumAromaticRings, NumHAcceptors, NumHDonors, NumHeteroatoms, NumRadicalElectrons, NumSaturatedCarbocycles, NumSaturatedHeterocycles, NumSaturatedRings, RingCount

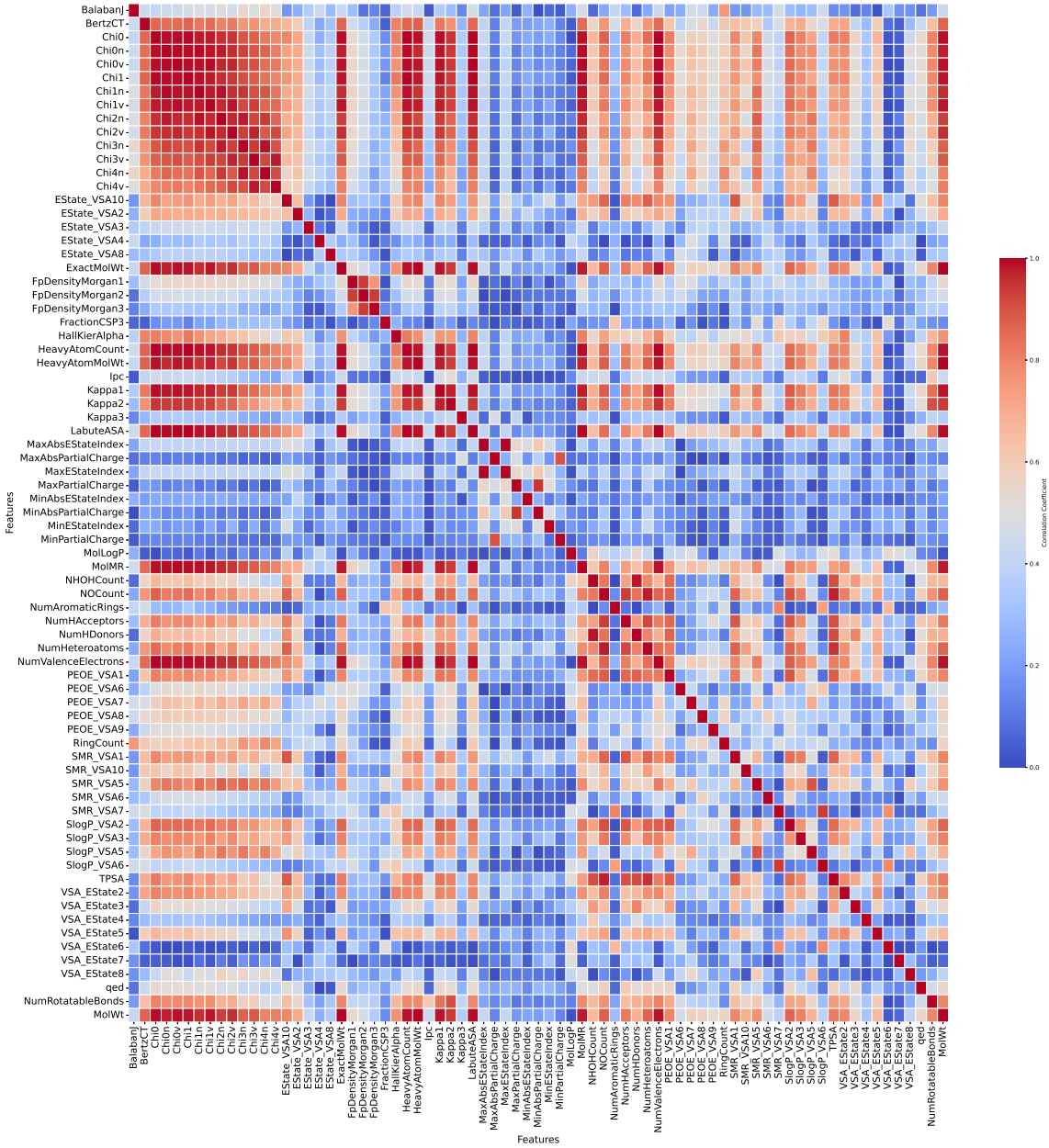

Figure S1: Heatmap of Pearson Correlation coefficient of pair-wise descriptors.

ATC code: To illustrate the construction of the feature matrix, we consider the drugs *Acebutolol Hydrochloride* and *Abemaciclib*. Each drug's ATC codes are broken down into their respective semantic set members, and the occurrences of each semantic set are counted to form the feature matrix. Here is the detailed example:

- *Acebutolol Hydrochloride*: ATC codes are $\{C07AB04, C07BB04\}$,
- *Abemaciclib*: ATC code is $\{L01EF03\}$

The semantic set members for these ATC codes are:

- C07AB04: $\{C, C07, C07A, C07AB, C07AB04\}$,
- C07BB04: $\{C, C07, C07B, C07BB, C07BB04\}$,
- and L01EF03: $\{L, L01, L01E, L01EF, L01EF03\}$.

The feature matrix based on the semantic sets of these ATC codes is shown in Table S4.

Table S4: Construction of ATC code feature matrix for drug-ATC code pairs.

Drug Name	C	C07	C07A	C07B	C07AB	C07BB	C07AB04	C07BB04	L	L01	L01E	L01EF	L01EF03
Acebutolol Hydrochloride	2	2	1	1	1	1	1	1	0	0	0	0	0
Abemaciclib	0	0	0	0	0	0	0	0	1	1	1	1	1

Parameter settings: The parameter settings for both feature matrices from the training dataset are optimized through five-fold cross-validation and grid search, with the optimal parameters chosen based on accuracy. The results of FM and CFM are summarized in Tables S5 and S6, respectively. Most of the parametric values are taken as 5 equally spaced values between a particular range.

Table S5: Parameter settings of different machine learning algorithms using FM.

Algorithm	Parameters	Parameter settings	Best values	Accuracy
RF	n_estimators	[100, 500]	500	0.7872
	min_samples_leaf	[1, 20]	1	
SVM	c	[0.03125, 32768]	32	0.7831
	gamma	[0.00003051757, 16]	0.5946	
	kernel	rbf, linear, poly, and sigmoid	rbf	
KNN	k	[3, 30]	23	0.7703
DT	max_depth	[10, 40]	10	0.7304
	min_samples_split	[2, 20]	10	
	min_samples_leaf	[1, 20]	11	
LR	c	[0.1, 10]	0.1	0.7676
	solver	liblinear, saga	liblinear	
GB	learning_rate	[0.05, 0.2]	0.2	0.7851
	n_estimators	[40, 70]	70	
	max_depth	[4, 8]	8	
AdaBoost	n_estimators	[50, 300]	300	0.7716
	learning_rate	[0.01, 1.0]	0.2575	
EGB	n_estimators	[100, 500]	500	0.7973
	learning_rate	[0.01, 0.2]	0.105	
	max_depth	[3, 10]	6	
	subsample	[0.5, 1.0]	0.5	
	colsample_bytree	[0.5, 1.0]	0.625	
CatBoost	iterations	[100, 500]	100	0.7932
	learning_rate	[0.001, 0.1]	0.0505	
	depth	[1, 10]	10	
BioGCN	size of hidden layer 1	[32, 64, 128, 256]	128	0.7689
	size of hidden layer 2	[16, 32, 64]	64	
	dropout rate	[0.1, 0.2, 0.3, 0.4, 0.5]	0.3	
	learning rate	[0.01, 0.001, 0.0001]	0.01	

Table S6: Parameter settings of different machine learning algorithms using CFM.

Algorithm	Parameters	Parameter settings	Best values	Accuracy
RF	n_estimators	[100, 500]	500	0.7797
	min_samples_leaf	[1, 20]	1	
SVM	c	[0.03125, 32768]	1024	0.7716
	gamma	[0.00003051757, 16]	2.2097×10^{-2}	
	kernel	rbf, linear, poly, and sigmoid	sigmoid	
KNN	k	[3, 30]	16	0.7649
DT	max_depth	[10, 40]	10	0.7264
	min_samples_split	[2, 20]	20	
	min_samples_leaf	[1, 20]	1	
LR	c	[0.1, 10]	2.575	0.7865
	Solver	liblinear, saga	liblinear	
GB	learning_rate	[0.05, 0.2]	0.125	0.7824
	n_estimators	[40, 70]	47	
	max_depth	[4, 8]	4	
AdaBoost	n_estimators	[50, 300]	300	0.7743
	learning_rate	[0.01, 1.0]	0.505	
EGB	n_estimators	[100, 500]	300	0.8014
	learning_rate	[0.01, 0.2]	0.1525	
	max_depth	[3, 10]	6	
	subsample	[0.5, 1.0]	0.5	
	colsample_bytree	[0.5, 1.0]	0.625	
CatBoost	iterations	[100, 500]	200	0.7939
	learning_rate	[0.001, 0.1]	0.1	
	depth	[1, 10]	5	
BioGCN	size of hidden layer 1	[32, 64, 128, 256]	32	0.7709
	size of hidden layer 2	[16, 32, 64]	16	
	dropout rate	[0.1, 0.2, 0.3, 0.4, 0.5]	0.5	
	learning rate	[0.01, 0.001, 0.0001]	0.0001	

Table S7: Comparison of machine learning model rankings based on six evaluation metrics on training data.

		RF	SVM	KNN	DT	LR	GB	NB	AdaBoost	EGB	CatBoost	BioGCN
None	Accuracy	2	9	7	10	5	4	11.0	6	3	1	8
	AUC	1	7	9	10	4	5	11	8	2	3	6
	Bcc	8	9	11	5	3	7	6	10	1	4	2
	Recall	2	7	1	10	6	3	11	4	8	5	9
	Precision	9	7	11	6	4	8	1	10	2	5	3
	F1 score	1	8	5	10	7	3	11	4	6	2	9
	Average	3.83	7.83	7.33	8.50	4.83	5	8.5	7	3.67	3.33	6.17
Undersampling	Accuracy	1	7	3	10	8	5	11	6	4	2	9
	AUC	1	6	9	10	5	4	11	8	3	2	7
	Bcc	1	6	9	10	5	4	11	8	3	2	7
	Recall	2	7	1	10	8	5	11	3	6	4	9
	Precision	1	5	9	11	4	6	10	8	2	3	7
	F1 score	1	7	2	10	8	6	11	4	5	3	9
	Average	1.17	6.33	5.50	10.17	6.33	5.00	10.83	6.17	3.83	2.67	8.00
Oversampling	Accuracy	1	4	10	9	5	6	11	8	2	3	7
	AUC	1	6	9	10	4	5	11	8	3	2	7
	Bcc	8	4	9	10	2	3	11	5	6	1	7
	Recall	1	4	10	8	6	5	11	9	2	3	7
	Precision	10	8	4	11	3	5	1	2	9	6	7
	F1 score	1	4	10	9	6	5	11	8	2	3	7
	Average	3.67	5.00	8.67	9.50	4.33	4.83	9.33	6.67	4.00	3.00	7.00
SMOTE	Accuracy	2	9	11	8	6	4	10	7	1	3	5
	AUC	1	8	9	10	5	4	11	7	2	3	6
	Bcc	4	8	9	10	2	5	11	6	1	3	7
	Recall	1	9	11	8	6	4	10	7	3	2	5
	Precision	10	4	1	11	3	7	2	5	6	8	9
	F1 score	1	9	11	8	6	4	10	7	2	3	5
	Average	3.17	7.83	8.67	9.17	4.67	4.67	9.00	6.50	2.50	3.67	6.17

Table S8: Comparison of machine learning model rankings based on six evaluation metrics on test dataset 1.

		RF	SVM	KNN	DT	LR	GB	NB	AdaBoost	EGB	CatBoost	BioGCN
None	Accuracy	10	3	2	8	1	7	6	9	5	11	4
	AUC	7	3	4	11	2	5	6	10	9	8	1
	Bcc	7	3	9	11	2	5	4	10	6	8	1
	Recall	10	3	1	7	2	8	6	9	5	11	4
	Precision	5	6	10	11	7	1	3	9	8	4	2
	F1 score	10	3	2	7	1	8	6	9	5	11	4
	Average	8.17	3.50	4.67	9.17	2.50	5.67	5.17	9.33	6.33	8.83	2.67
Undersampling	Accuracy	8	1	3	4	2	10	7	11	5	6	9
	AUC	6	1	4	10	2	7	8	11	9	5	3
	Bcc	7	1	3	10	2	8	4	11	9	5	6
	Recall	8	1	4	3	2	10	7	11	5	6	9
	Precision	4	8	7	11	9	2	5	1	10	6	3
	F1 score	8	1	4	3	2	10	7	11	5	6	9
	Average	6.83	2.17	4.17	6.83	3.17	7.83	6.33	9.33	7.17	5.67	6.50
Oversampling	Accuracy	9	1	4	5	3	10	8	11	6	7	2
	AUC	6	1	4	10	2	7	8	11	9	5	3
	Bcc	7	1	4	10	3	8	5	11	9	6	2
	Recall	9	1	5	4	3	10	8	11	6	7	2
	Precision	3	8	6	11	9	2	4	1	10	5	7
	F1 score	9	1	5	4	3	10	8	11	6	7	2
	Average	7.17	2.17	4.67	7.33	3.83	7.83	6.83	9.33	7.67	6.17	3.00
SMOTE	Accuracy	8	3	5	10	2	9	4	11	6	7	1
	AUC	4	3	6	11	1	9	7	10	8	5	2
	Bcc	6	7	4	10	2	9	1	11	8	5	3
	Recall	8	2	6	9	2	10	4	11	5	7	1
	Precision	1	9	4	10	7	2	3	11	6	5	8
	F1 score	8	3	6	9	2	10	4	11	5	7	1
	Average	5.83	4.50	5.17	9.83	2.67	8.17	3.83	10.83	6.33	6.00	2.67

Table S9: Comparison of machine learning model rankings based on six evaluation metrics on test dataset 2.

		RF	SVM	KNN	DT	LR	GB	NB	AdaBoost	EGB	CatBoost	BioGCN
None	Accuracy	7	3	3	11	1	9	5	8	5	10	2
	AUC	6	3	7	11	2	9	4	5	10	8	1
	Bcc	5	4	8	11	2	6	3	7	9	10	1
	Recall	6	4	1	9	1	9	9	6	3	8	5
	Precision	6	4	9	11	5	3	1	7	8	10	2
	F1 score	6	5	2	11	1	10	8	7	4	9	3
	Average	6.00	3.83	5.00	10.67	2.00	7.67	5.00	6.67	6.50	9.17	2.33
Undersampling	Accuracy	11	1	3	7	3	10	6	8	5	9	2
	AUC	4	1	5	11	3	9	10	6	7	8	2
	Bcc	11	1	5	8	3	10	4	7	6	9	2
	Recall	11	1	2	6	4	10	7	9	5	8	3
	Precision	1	6	10	6	5	6	1	1	4	10	9
	F1 score	11	1	3	6	4	10	7	9	5	8	2
	Average	8.17	1.83	4.67	7.33	3.67	9.17	5.83	6.67	5.33	8.67	3.33
Oversampling	Accuracy	4	2	10	4	2	6	6	11	9	6	1
	AUC	4	2	10	7	3	6	5	9	11	8	1
	Bcc	6	2	10	6	2	5	4	9	11	8	1
	Recall	4	2	9	4	2	8	9	11	4	7	1
	Precision	7	4	10	7	4	3	2	1	11	9	6
	F1 score	4	2	10	4	2	8	9	11	7	6	1
	Average	4.83	2.33	9.83	5.33	2.50	6.00	5.83	8.67	8.83	7.33	1.83
SMOTE	Accuracy	3	6	11	7	2	4	5	10	8	8	1
	AUC	3	5	7	6	2	8	4	9	11	10	1
	Bcc	3	6	11	7	2	5	4	9	10	8	1
	Recall	3	3	10	7	2	5	7	11	6	9	1
	Precision	3	9	10	7	5	4	1	2	10	8	5
	F1 score	3	5	10	7	2	4	6	11	8	9	1
	Average	3.00	5.67	9.83	6.83	2.50	5.00	4.50	8.67	8.83	8.67	1.67