

Volgograd State Medical University

Research Center of Innovative Medicines

*Laboratory for Information Technology in Pharmacology
and Computer Modeling of Drugs*



**CONSENSUS MODELING
OF ANXIOLYTIC ACTIVITY
OF CHEMICAL COMPOUNDS
BY CONVOLUTIONAL NEURAL
NETWORKS**

Vassiliev Pavel Mikhailovich

Research Center of Innovative Medicines



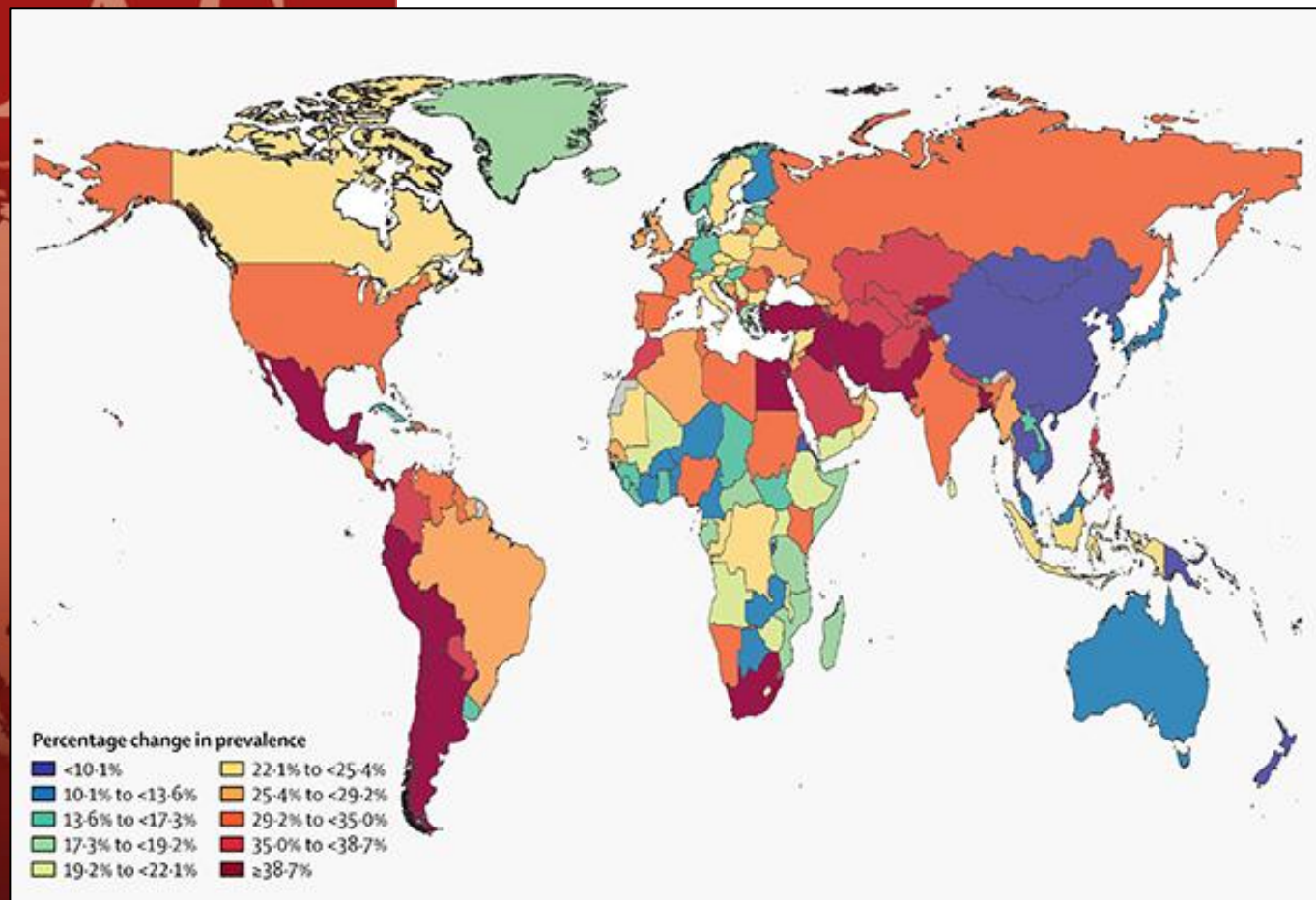
Depression and anxiety disorders

MENTAL HEALTH

ATLAS

2020

635 million people



Build classification consensus ensemble model of dependence of anxiolytic activity of chemical compounds from their structural QL descriptors using method of convolution artificial neural networks

Tasks

- **Formation of verified training set for known anxiolytic substances**
- **Calculation of QL descriptor representation for training set compounds**
- **Line convolution of find structural parameters**
- **Training of neural networks and formation of consensus ensemble model**
- **Prediction of anxiolytic activity for new compounds**

Finding information about known compounds

ChEMBL Search in ChEMBL

EBI > Databases > Chemical Biology > ChEMBL Database > Search Results

Search Results

All Results 3717284 Compounds 2157379 Targets 14855 Assays 1458215 Documents 84092 Cells 1991 Tissues 752

Compounds

Anxiolytic activity

1 939 entries

Show Full Query ?

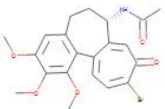
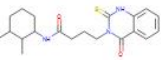
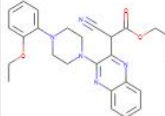
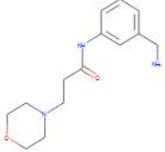
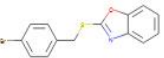
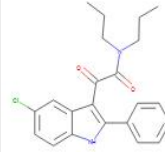
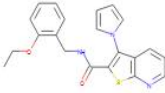
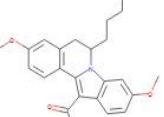

2,157,379 Compounds
0 Selected - Select All
Browse Activities ?

Table Cards Graph Heatmap

CSV TSV SDF

Records per page: 24 Select All

Showing 1-24 out of 2,157,379 records

 <p>CHEMBL166302 Name: No Data Max Phase: 0 Full Mw: 448.31 Alogp: 3.63</p>	 <p>CHEMBL1542218 Name: No Data Max Phase: 0 Full Mw: 373.52 Alogp: 3.78</p>	 <p>CHEMBL1454107 Name: No Data Max Phase: 0 Full Mw: 457.53 Alogp: 3.69</p>
 <p>CHEMBL3465724 Name: No Data Max Phase: 0 Full Mw: 263.34 Alogp: 0.81</p>	 <p>CHEMBL484491 Name: No Data Max Phase: 0 Full Mw: 320.21 Alogp: 4.88</p>	 <p>CHEMBL54627 Name: No Data Max Phase: 0 Full Mw: 382.89 Alogp: 5.32</p>
 <p>CHEMBL1316827 Name: No Data Max Phase: 0 Full Mw: 377.47 Alogp: 4.42</p>	 <p>CHEMBL120536 Name: No Data Max Phase: 0 Full Mw: 377.48 Alogp: 5.82</p>	 <p>CHEMBL356790 Name: No Data Max Phase: 0 Full Mw: 430.64 Alogp: 5.63</p>

6

Unification of techniques and activity clustering

Unification and aggregation of techniques

Anxiolytic — 707

Clustering Perc.stw*

- Cluster Analysis (Data for Clustering CHEMBL25 Glycation v01)
 - K-means clustering results dialog
 - Cluster Means (Data for Clustering CHEMBL25 Glycation)

Variable	Cluster No. 1	Cluster No. 2	Cluster No. 3
Inhib	63.63278	93.15059	29.47273
 - Euclidean Distances between Clusters (Data for Clustering CHEMBL25 Glycation)

Cluster Number	No. 1	No. 2	No. 3
No. 1	0.00000	871.3011	1166.909
No. 2	29.51781	0.0000	4054.870
No. 3	34.16005	63.6779	0.000
 - Analysis of Variance (Data for Clustering CHEMBL25 Glycation)

Variable	No. 1	No. 2	No. 3
log(1/IC50)	3.376441	4.259640	5.064370
 - Descriptive Statistics for Cluster 1 (Data for Clustering CHEMBL25 Glycation)

Variable	Mean	Standard Deviation	Variance
log(1/IC50)	3.376441	0.201177	0.040472
 - Descriptive Statistics for Cluster 2 (Data for Clustering CHEMBL25 Glycation)

Variable	Mean	Standard Deviation	Variance
log(1/IC50)	4.259640	0.222052	0.049307
 - Descriptive Statistics for Cluster 3 (Data for Clustering CHEMBL25 Glycation)

Variable	Mean	Standard Deviation	Variance
log(1/IC50)	5.064370	0.203475	0.041402
 - Members of Cluster Number 1 (Data for Clustering CHEMBL25 Glycation)

Variable	Distance
361.000000	0.412706
192.000000	0.375586
388.000000	0.371192
377.000000	0.352345
 - Members of Cluster Number 2 (Data for Clustering CHEMBL25 Glycation)

Variable	Distance
348.000000	0.395968
108.000000	0.361599
144.000000	0.357083
52.000000	0.352970
 - Members of Cluster Number 3 (Data for Clustering CHEMBL25 Glycation)

Variable	Distance
406.000000	
387.000000	
264.000000	
389.000000	

Clustering by k-means method

4 activity classes

High, Moderate, Low, Inactive

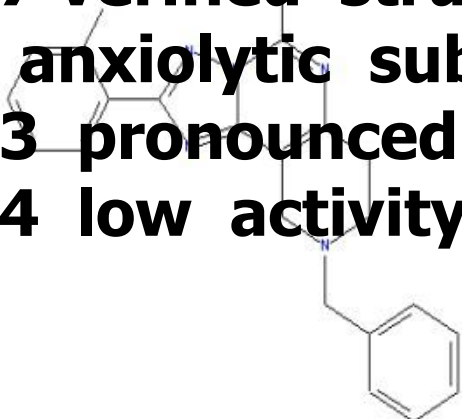
Formation of verified databases

Anxiolytic Substances v11 - Corrected

24 January 2022

Structure

**537 verified structures
of anxiolytic substances:
273 pronounced activity
264 low activity**



Activity

Standard Type

ED50

pChEMBL Value

Assay

Assay ChEMBL ID

CHEMBL779993

Assay Description

Anxiolytic response measurement in rats.

Mol_ID: 1
Brutto formula: C₂₁H₁₈FN₅O
Mol weight: 375.3
Salt Component:

Compound Codes & Names

MolFileName

CHEMBL99259

CODE - Molecule ChEMBL ID

CHEMBL99259

Molecule Name

Compound Key

8g

References

Document_ChEMBL_ID

CHEMBL1125833

Document Journal

J. Med. Chem.

Document_Year

1991

Activity Processing

Gen_Assay_IDs

Rat-07a

Gen_Assay_ChEMBLID

CHEMBL3256343 CHEMBL784521 CHEMBL84

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации базы данных

№ 2022621744

Соединения с анксиолитической активностью

Правообладатель: Федеральное государственное бюджетное образовательное учреждение высшего образования «Волгоградский государственный медицинский университет» Министерства здравоохранения Российской Федерации (RU)

Авторы: Васильев Павел Михайлович (RU), Мальцев Дмитрий Васильевич (RU), Перфильев Максим Алексеевич (RU), Спасов Александр Алексеевич (RU), Скрипка Мария Олеговна (RU), Кочетков Андрей Николаевич (RU)

Заявка № 2022621682

Дата поступления 13 июля 2022 г.

Дата государственной регистрации в Реестре баз данных 15 июля 2022 г.

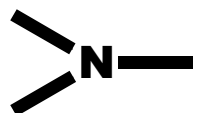
Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов



Elementary descriptors of QL language

Structural descriptors



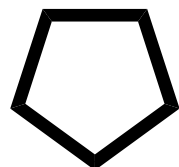
-N<

heteroatomic



-CH3

carbonic



Cyc05

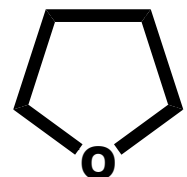
cyclic

Length descriptors



03

path length



-1

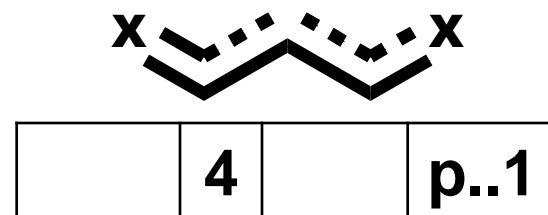
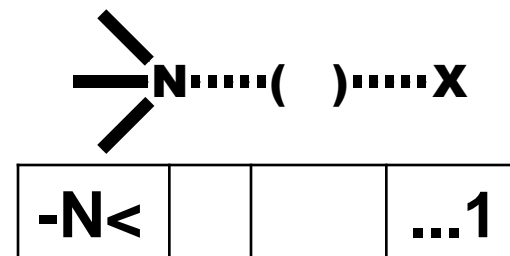
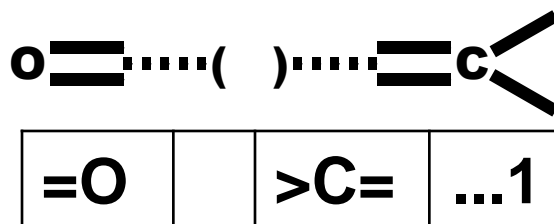
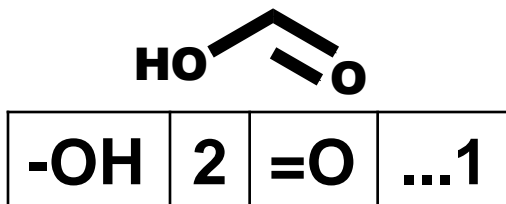
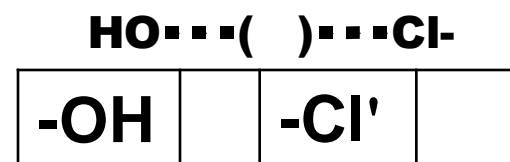
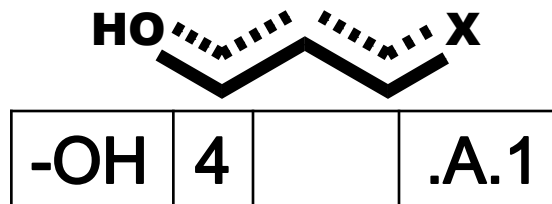
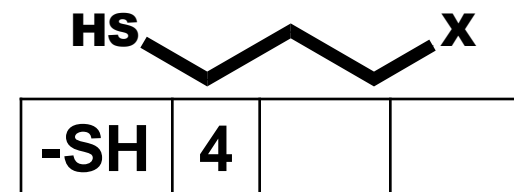
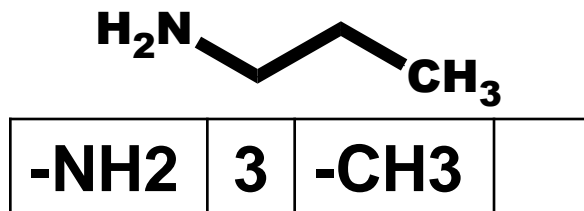
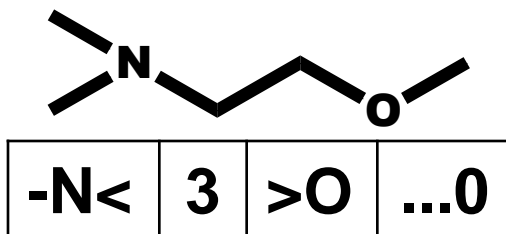
inclusion index



p..1

Bond descriptors

Composite descriptors of QL language



**4 rank
Basic**

3 rank

2 rank

Calculation of QL descriptors

Трансляция набора структур в формат QL-2

О программе

Набор структур: LPS_v02.sdf

Выходной файл QL: LPS_v02

Транслируется набор структур: LPS_v02

23%

8686 species of QL descriptors of 11 different types

Создание набора файлов для прогноза

Входной файл (QL): LPS_v02

Выходной набор: LPS_v02

Создать набор

Идет создание набора файлов для прогноза...

9%

Correlation of one sequence with the reversed second

$$y(n) = \sum_{m=0}^n h(n-m)x(m)$$

8686 QL descriptors \longrightarrow 66 convolution variables

Example of QL descriptor convolution

SD	-NH2	>NH	-N<	-N=	#N	>N+=	-OH	>O	=O
	0	0	2	3	0	0	1	0	0
BD	...0	...1	..n1	.a.0	.a.1	.A.0	.A.1	p..0	p..1
	12	30	0	10	23	15	30	0	0

Initial data

SD	-NH2	>NH	-N<	-N=	#N	>N+=	-OH	>O	=O
	0	0	2	3	0	0	1	0	0
BD	p..1	p..0	.A.1	.A.0	.a.1	.a.0	..n1	...1	...0
	0	0	30	15	23	10	0	30	12

Data for convolution

$$\begin{aligned}
 \mathbf{SD \cdot BD} &= (0 \cdot 0) + (0 \cdot 0) + (2 \cdot 30) + (3 \cdot 15) + (0 \cdot 23) + (0 \cdot 10) + \\
 &(1 \cdot 0) + (0 \cdot 30) + (0 \cdot 12) = 0 + 0 + 60 + 45 + 0 + 0 + 0 + 0 + 0 \\
 &= \mathbf{105}
 \end{aligned}$$

Formation of training set

Mol_ID	Code	LevHM	SD_1-1	SD_1-2	SD_1-3	SD_1-4	SD_1-5	SD_1-6	SD_1-7	SD_1-8	SD_1-9	SD_1-10	SD_1-11	SD_2-2	SD_2-3	SD_2-4
1	CHEMBL101045	hm	1089	3003	3003	6006	3003	6006	3003	3003	6006	3003	3003	8281	8281	16562
2	CHEMBL101304	hm	676	2418	2418	4836	2418	4836	2418	2418	4836	2418	2418	8649	8649	17298
3	CHEMBL13662	hm	529	3266	3266	6532	3266	6532	3266	3266	6532	3266	3266	20164	20164	40328
4	CHEMBL153823	hm	361	1615	1615	3230	1615	3230	1615	1615	3230	1615	1615	7225	7225	14450
5	CHEMBL154328	hm	2601	10455	10455	20910	10455	20910	10455	10455	20910	10455	10455	42025	42025	84050
6	CHEMBL154380	hm	2601	10455	10455	20910	10455	20910	10455	10455	20910	10455	10455	42025	42025	84050
7	CHEMBL155270	hm	361	1843	1843	3686	1843	3686	1843	1843	3686	1843	1843	9409	9409	18818
8	CHEMBL157522	hm	484	3036	3036	6072	3036	6072	3036	3036	6072	3036	3036	19044	19044	38088
9	CHEMBL187827	hm	529	2714	2714	5428	2714	5428	2714	2714	5428	2714	2714	13924	13924	27848
10	CHEMBL188330	hm	625	4500	4500	9000	4500	9000	4500	4500	9000	4500	4500	32400	32400	64800
11	CHEMBL189061	hm	784	7056	7056	14112	7056	14112	7056	7056	14112	7056	7056	63504	63504	127008
12	CHEMBL190772	hm	729	7101	7101	14202	7101	14202	7101	7101	14202	7101	7101	69169	69169	138338
13	CHEMBL20042	hm	625	3000	3000	6000	3000	6000	3000	3000	6000	3000	3000	14400	14400	28800
14	CHEMBL204240	hm	484	1276	1276	2552	1276	2552	1276	1276	2552	1276	1276	3364	3364	6728
15	CHEMBL2263292	hm	784	2044	2044	4088	2044	4088	2044	2044	4088	2044	2044	5329	5329	10658
16	CHEMBL2263293	hm	784	2044	2044	4088	2044	4088	2044	2044	4088	2044	2044	5329	5329	10658
17	CHEMBL2263296	hm	784	2604	2604	5208	2604	5208	2604	2604	5208	2604	2604	8649	8649	17298
18	CHEMBL239259	hm	1089	3960	3960	7920	3960	7920	3960	3960	7920	3960	3960	14400	14400	28800
19	CHEMBL260870	hm	1225	5845	5845	11690	5845	11690	5845	5845	11690	5845	5845	27889	27889	55778
20	CHEMBL260872	hm	1089	4224	4224	8448	4224	8448	4224	4224	8448	4224	4224	16384	16384	32768
21	CHEMBL260994	hm	961	3255	3255	6510	3255	6510	3255	3255	6510	3255	3255	11025	11025	22050
22	CHEMBL289742	hm	784	5348	5348	10696	5348	10696	5348	5348	10696	5348	5348	36481	36481	72962
23	CHEMBL290906	hm	784	5264	5264	10528	5264	10528	5264	5264	10528	5264	5264	35344	35344	70688
24	CHEMBL307116	hm	576	3552	3552	7104	3552	7104	3552	3552	7104	3552	3552	21904	21904	43808
25	CHEMBL3084529	hm	1444	7562	7562	15124	7562	15124	7562	7562	15124	7562	7562	39601	39601	79202
26	CHEMBL314608	hm	729	3186	3186	6372	3186	6372	3186	3186	6372	3186	3186	13924	13924	27848
27	CHEMBL319178	hm	1024	3488	3488	6976	3488	6976	3488	3488	6976	3488	3488	11881	11881	23762
28	CHEMBL319966	hm	625	2450	2450	4900	2450	4900	2450	2450	4900	2450	2450	9604	9604	19208
29	CHEMBL323519	hm	1296	3492	3492	6984	3492	6984	3492	3492	6984	3492	3492	9409	9409	18818
30	CHEMBL3246317	hm	729	3213	3213	6426	3213	6426	3213	3213	6426	3213	3213	14161	14161	28322
31	CHEMBL3251778	hm	841	3712	3712	7424	3712	7424	3712	3712	7424	3712	3712	16384	16384	32768
32	CHEMBL3277537	hm	784	3080	3080	6160	3080	6160	3080	3080	6160	3080	3080	12100	12100	24200
33	CHEMBL3277538	hm	961	3627	3627	7254	3627	7254	3627	3627	7254	3627	3627	13689	13689	27378
34	CHEMBL3277539	hm	1024	4320	4320	8640	4320	8640	4320	4320	8640	4320	4320	18225	18225	36450
35	CHEMBL3277542	hm	1024	3936	3936	7872	3936	7872	3936	3936	7872	3936	3936	15129	15129	30258
36	CHEMBL3277543	hm	729	2943	2943	5886	2943	5886	2943	2943	5886	2943	2943	11881	11881	23762
37	CHEMBL3277544	hm	784	3108	3108	6216	3108	6216	3108	3108	6216	3108	3108	12321	12321	24642

537 Anxiolytics

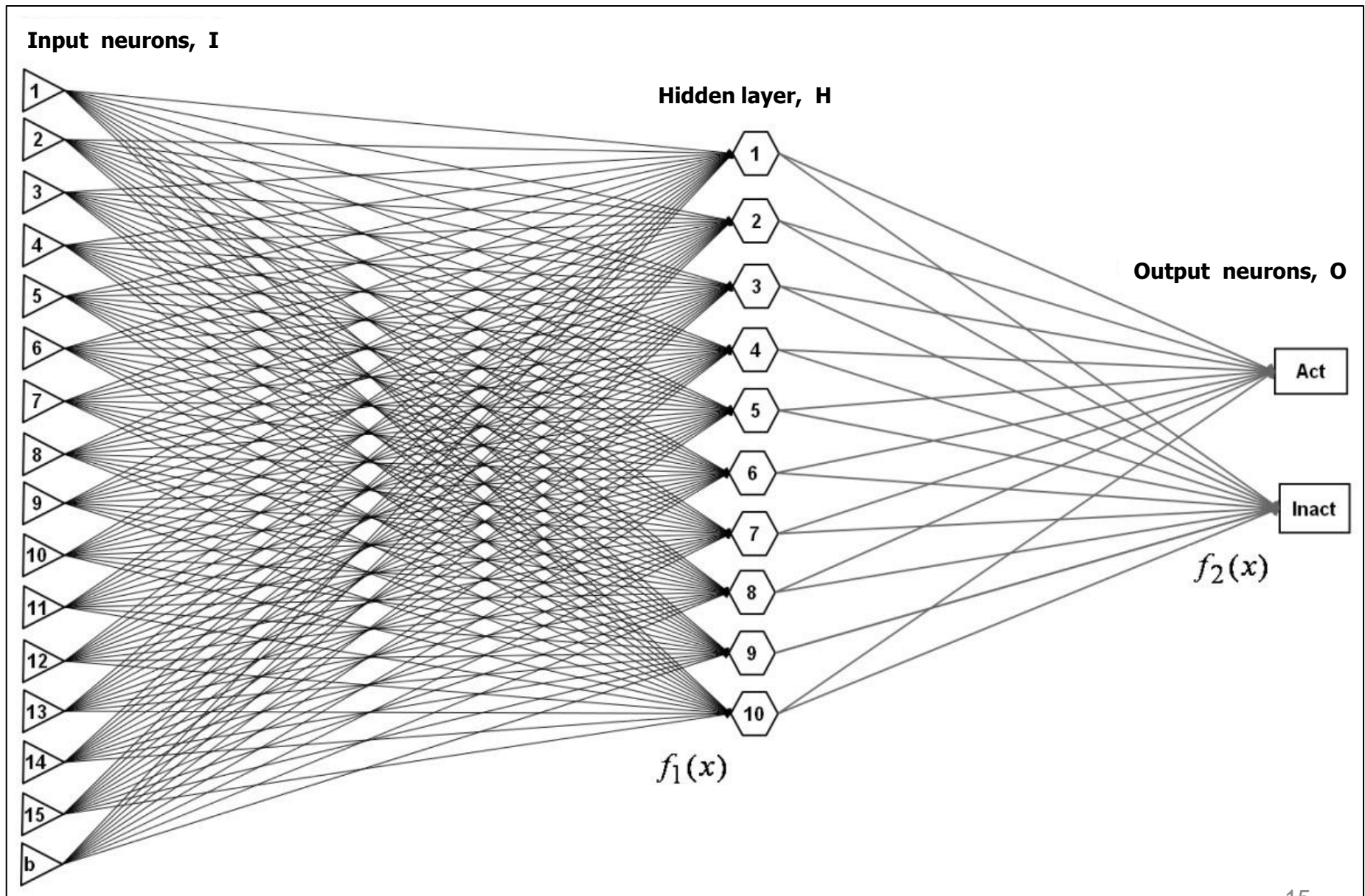
273 with pronounced activity

264 with low activity

66 neurons

7 samplings

Neural network with bottle-neck



Training of neural networks and formation of ensemble

7 sampling options

Net ...	Net name	Training ...	Test p...	Algorit...	Error fu...	Hidden ...	Output ...
1	MLP 22-1...	78.911565	80.555...	BFGS 14	SOS	Tanh	Logistic
2	MLP 22-1...	78.911565	80.555...	BFGS 14	SOS	Exponen...	Logistic
3	MLP 22-1...	79.591837	80.555...	BFGS 14	SOS	Exponen...	Tanh
4	MLP 22-1...	78.911565	80.555...	BFGS 14	SOS	Logistic	Logistic
5	MLP 22-9-2...	78.231233	80.555...	BFGS 11	SOS	Identity	Tanh
6	MLP 22-1...	78.911565	80.555...	BFGS 9	SOS	Tanh	Logistic

4000 trained neural networks for each sampling option

50 automatically selected neural networks

1 best neural network

Neural network training in progress...

Building network 145 (MLP 22-16-2, exp, logistic)
Cycle 40:
Classification rate: Train=85.034, Test=75

Ensemble of 7 neural networks

~30 000 networks were trained

Options

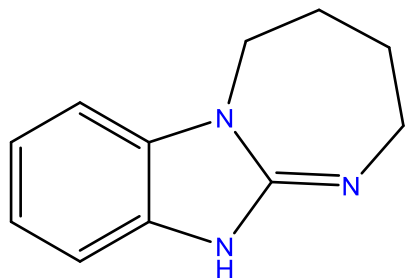
- Train
- Test
- Validation
- Missing

Accuracy of neural network ensemble

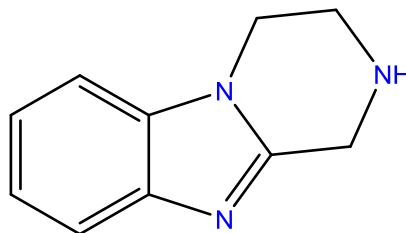
1	Code	Метка	Sampl1	Sampl2	Sampl3	Sampl4	Sampl5	Sampl6	Sampl7	Ind1	Ind2	Ind3	Ind4	Ind5	Ind6	Ind7	Cons>=4	Cons>=5	Cons>=6	Cons=7	GenCons	Correct
2	CHEMBL100144	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
3	CHEMBL101045	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
4	CHEMBL101304	hm	hm	hm	hm	nhm	hm	hm	hm	1	1	1		1	1	1	hm	hm	hm		hm	1
5	CHEMBL1059	hm	hm	hm	hm	nhm	nhm	hm	hm	1	1	1			1	1	hm	hm			hm	1
6	CHEMBL1112	hm	hm	hm	nhm	hm	hm	hm	hm	1	1		1	1	1	1	hm	hm	hm		hm	1
7	CHEMBL13662	hm	hm	hm	nhm	nhm	hm	hm	hm	1	1			1	1	1	hm	hm			hm	1
8	CHEMBL153823	hm	hm	hm	hm	nhm	nhm	hm	hm	1	1	1		1	1	1	hm	hm			hm	1
9	CHEMBL153953	hm	nhm	hm	hm	nhm	nhm	hm	hm	1	1	1		1	1	1	hm	hm			hm	1
10	CHEMBL154043	hm	hm	hm	hm	nhm	nhm	hm	hm	1	1	1		1	1	1	hm	hm	hm		hm	1
11	CHEMBL154328	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
12	CHEMBL154380	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
13	CHEMBL154754	hm	hm	hm	hm	nhm	hm	hm	hm	1	1	1		1	1	1	hm	hm	hm		hm	1
14	CHEMBL155270	hm	nhm	hm	hm	nhm	nhm	hm	hm	1	1	1		1	1	1	hm	hm			hm	1
15	CHEMBL157522	hm	hm	hm	nhm	nhm	nhm	hm	hm	1	1	1		1	1	1	hm				hm	1
16	CHEMBL1836783	hm	hm	hm	hm	hm	hm	nhm	hm	1	1	1	1	1		1	hm	hm	hm		hm	1
17	CHEMBL187827	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
18	CHEMBL188138	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
19	CHEMBL188280	hm	nhm	hm	nhm	nhm	nhm	nhm	nhm													nhm
20	CHEMBL188335	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
21	CHEMBL189061	hm	hm	hm	hm	hm	nhm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm		hm	1
22	CHEMBL190772	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
23	CHEMBL1949963	hm	hm	hm	hm	hm	hm	nhm	hm	1	1	1	1	1		1	hm	hm	hm		hm	1
24	CHEMBL20042	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
25	CHEMBL204040	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
26	CHEMBL204091	hm	nhm	nhm	nhm	nhm	hm	nhm	nhm					1								nhm
27	CHEMBL204240	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
28	CHEMBL204755	hm	nhm	nhm	nhm	nhm	nhm	nhm	nhm													nhm
29	CHEMBL207493	hm	hm	hm	hm	nhm	nhm	hm	hm	1	1	1			1	1	hm	hm			hm	1
30	CHEMBL207517	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
31	CHEMBL207678	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
32	CHEMBL2263292	hm	nhm	hm	hm	hm	hm	hm	nhm		1	1	1	1	1		hm	hm			hm	1
33	CHEMBL2263293	hm	nhm	hm	hm	hm	hm	hm	nhm		1	1	1	1	1		hm	hm			hm	1
34	CHEMBL2263294	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
35	CHEMBL2263295	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
36	CHEMBL2263296	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
37	CHEMBL2263297	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
38	CHEMBL2263298	hm	hm	hm	hm	hm	hm	hm	hm	1	1	1	1	1	1	1	hm	hm	hm	hm	hm	1
39	CHEMBL231825	hm	hm	hm	nhm	hm	hm	nhm	hm	1	1		1	1		1	hm	hm			hm	1
40	CHEMBL232188	hm	hm	hm	nhm	nhm	hm	hm	hm	1	1			1	1	1	hm	hm			hm	1

Accuracy $F_0 = 78\%$
Sensitivity $F_a = 85\%$
Specificity $F_n = 70\%$

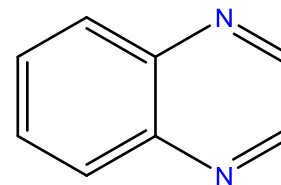
Prediction results



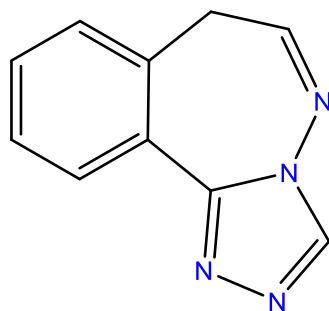
Diazepinobenzimidazole



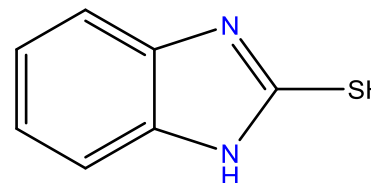
Benzoimidazopyrazine



Quinoxaline



Triazolobenzodiazepine



Mercaptobenzimidazole

97 new compounds – 53 active

Laboratory for Information Technology in Pharmacology and Computer Modeling of Drugs



~40 Tflops

A photograph of the Statue of Liberty at night, illuminated against a dark blue sky with scattered clouds. The statue stands on a grassy hill, with its torch held high. The scene is reflected in a body of water in the foreground, which is lit by warm yellow lights along the shore. The overall atmosphere is serene and majestic.

Thank You for Your
attention!