

©Коллектив авторов

## МУЛЬТИТАРГЕТНАЯ НЕЙРОСЕТЕВАЯ МОДЕЛЬ АНКСИОЛИТИЧЕСКОЙ АКТИВНОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ НА ОСНОВЕ КОРРЕЛЯЦИОННОЙ СВЁРТКИ СПЕКТРОВ ЭНЕРГИЙ МНОЖЕСТВЕННОГО ДОКИНГА

П.М. Васильев\*, М.А. Перфильев, А.В. Голубева, А.Н. Кочетков, Д.В. Мальцев

Волгоградский государственный медицинский университет,  
400131, Волгоград, пл. Павших борцов, 1; \*эл. почта: pvassiliev@mail.ru

Тревожные расстройства являются одной из самых распространённых в мире патологий психического здоровья, которые требуют поиска и создания новых эффективных фармакологически активных веществ. В связи с этим разработка с использованием методов искусственного интеллекта новых подходов к поиску анксиолитических веществ является актуальным направлением современной биоинформатики и фармакологии. В настоящей работе методом искусственных нейронных сетей построена мультитаргетная модель зависимости анксиолитической активности химических соединений от их интегральной аффинности к релевантным белкам-мишеням, основанная на корреляционной свёртке спектров энергии множественного докинга. Для этого на основе ранее созданной базы данных была сформирована обучающая выборка по структуре и активности 537 известных анксиолитических веществ и построены оптимизированные 3D-модели этих соединений. Выявлены 22 биомишени, предположительно релевантные анксиолитической активности, и найдены их валидные 3D-модели. Для каждой такой биомишени по всему её объёму сформированы 27 пространств для множественного докинга. Выполнен множественный ансамблевый молекулярный докинг 537 известных анксиолитических соединений во все пространства релевантных белков-мишеней. Проведена корреляционная свёртка рассчитанных спектров энергий множественного докинга. С использованием семи вариантов обучения на основе искусственных многослойных перцептронных нейронных сетей построена мультитаргетная модель зависимости анксиолитической активности химических соединений от 22 параметров корреляционной свёртки спектров энергий их множественного докинга. Выполнена оценка прогностической способности созданной модели, общая точность которой составила  $\text{Acc} = 91,2\%$  и  $\text{AUC}_{\text{ROC}} = 94,4\%$ , при статистической достоверности  $p < 1 \times 10^{-15}$ . Найденная модель используется в поиске новых веществ с высокой анксиолитической активностью.

**Ключевые слова:** анксиолитическая активность; мультитаргетная аффинность; множественный молекулярный докинг; спектр энергий докинга; корреляционная свёртка; искусственные нейронные сети

**DOI:** 10.18097/PBMC20247006428

### ВВЕДЕНИЕ

По данным ВОЗ за 2023 г. [1], в мире почти 1 млрд. человек страдают психическими расстройствами, из них основную часть составляют тревожные расстройства. По данным Росстата за 2023 г. [2], в России насчитывается примерно 3,5 млн. человек, страдающих тревожными расстройствами. Тревожные расстройства, характеризующиеся чрезмерным беспокойством, страхом и избегающим поведением, осложняют повседневную жизнь пациентов и ограничивают их социальную активность и профессиональные возможности. Всё это делает актуальным поиск новых анксиолитических препаратов.

Методы машинного обучения давно и успешно применяются в поиске *in silico* соединений с фармакологической активностью [3], причём особенно эффективно в этом отношении показала себя технология искусственных нейронных сетей (собираательно называемая методами искусственного интеллекта) [4]. Однако в PubMed нам не удалось найти ни одной публикации, посвящённой поиску *in silico* с использованием методов искусственного интеллекта веществ с анксиолитической активностью. Вероятно, это связано с тем, что анксиолитические эффекты носят сложный системный мультитаргетный

характер и обусловлены воздействием соединений на совокупность достаточно большого числа релевантных биомишеней.

Таким образом, разработка с использованием методов искусственного интеллекта новых подходов к поиску анксиолитических веществ является одним из актуальных направлений современной биоинформатики и фармакологии.

Ранее в нашей работе [5] было показано, что спектр энергий докинга химических соединений в множество пространств релевантного белка является намного более достоверной метрикой аффинности лигандов к биомишеням, в сравнении с единичной энергией их докинга в специфический сайт. С учётом этого, нами была построена основанная на множественном докинге нейросетевая модель зависимости анксиолитической активности химических соединений от спектра энергий их множественного докинга в ГАМК<sub>A</sub>-рецептор, основанная на учёте интегральной аффинности прогнозируемых соединений только к одной биомишени [6].

Целью настоящей работы была разработка более универсальной мультитаргетной нейросетевой модели зависимости анксиолитической активности химических соединений от параметров корреляционной свёртки спектров энергий их множественного докинга.

Для достижения поставленной цели необходимо было решить следующие задачи.

1. Сформировать обучающую выборку по структуре и уровню анксиолитической активности известных соединений.
2. Построить оптимизированные 3D-модели этих соединений.
3. Выявить биомишени, предположительно релевантные анксиолитической активности.
4. Найти валидные 3D-модели белков-мишеней, релевантных анксиолитической активности.
5. Сформировать пространства для множественного докинга по всему объёму каждой валидированной 3D-модели релевантных белков-мишеней.
6. Выполнить множественный ансамблевый молекулярный докинг известных анксиолитических соединений во все пространства всех валидированных 3D-моделей релевантных белков-мишеней.
7. Провести корреляционную свёртку рассчитанных спектров энергий множественного докинга.
8. Выполнить обучение нейронных сетей на полученных свёрточных переменных и сформировать мультитаргетную нейросетевую модель зависимости анксиолитической активности химических соединений от параметров корреляционной свёртки спектров энергий их множественного докинга.

## МЕТОДИКА

### *Обучающая выборка по структуре и уровню анксиолитической активности известных соединений*

Обучающая выборка была сформирована на основе оригинальной верифицированной базы данных [7] по структуре и анксиолитической активности 537 известных экспериментально изученных веществ с проверенными структурными формулами и унифицированным уровнем анксиолитической активности. В неё вошли 273 соединения с выраженной анксиолитической активностью и 264 слабо активных или неактивных соединения.

### *Оптимизированные 3D-модели соединений*

Оптимизированные 3D-модели 537 соединений были построены последовательно методом молекулярной механики в программе MarvinSketch 15.6.15 [8] и затем полуэмпирическим квантово-химическим методом PM7 в программе MOPAC2012 [9]. Сначала отдельно для каждого соединения в MarvinSketch были построены по 10 конформеров с наименьшей энергией. Затем все построенные конформеры были оптимизированы в MOPAC2012 и для каждого соединения среди его оптимизированных конформеров был отобран один конформер с наименьшей общей энергией. Всего по данной схеме было обработано 10749 конформеров, в результате построено 537 оптимизированных 3D-моделей соединений обучающей выборки.

### *Биомишени, релевантные анксиолитической активности*

В системе Open Targets [10] был получен список 2057 биомишеней, ассоциированных с тревожными расстройствами по данным текст-майнинга имеющихся публикаций. Из оригинальной QSAR-базы системы Microcosm BioS 20.6.6 [11] взят список 2697 биомишеней человека с экспериментальными данными по различным видам таргетной активности. Пересечением этих двух списков был получен список из 92 биомишеней человека, предположительно релевантных анксиолитической активности и имеющих достоверное подтверждение в виде экспериментальных данных по активности соединений. С использованием 273 соединений из оригинальной базы данных [7], проявляющих анксиолитическую активность, и оригинальных систем IT Microcosm 7.3 [12] и Microcosm BioS 20.6.6 [11], а также метода структурного сходства с применением QL-модифицированного коэффициента Танимото для каждой биомишени были найдены по 10 наиболее близких по структуре известных соединений, экспериментально изученных на этот вид таргетной активности. По полученным данным были вычислены 92 средних показателя уровня таргетной активности Ind, интервал изменений от Ind = +5 (очень высокая) до Ind = -5 (неактивно); Ind = 0 соответствует средней активности. В качестве наиболее вероятных релевантных выраженому уровню анксиолитической активности отобраны 22 биомишени, имеющие значения Ind  $\geq 1$ .

### *Валидные 3D-модели релевантных биомишеней*

Для 22 возможных релевантных биомишеней из найденных в базах данных PDBe (<https://www.ebi.ac.uk/pdbe/>) и RCSB PDB (<https://www.rcsb.org/>) рентгеноструктурных 3D-моделей были отобраны 277 наиболее качественных 3D-моделей. Критериями качества служили: 1) максимальная длина смоделированной аминокислотной последовательности; 2) высокое разрешение; 3) минимальное число фрагментов. Среди этих 277 3D-моделей по методике, описанной в работе [13], были выявлены 22 валидных 3D-модели, по одной для каждой биомишени.

### *Пространства для множественного молекулярного докинга*

На каждой валидной 3D-модели каждой релевантной биомишени с помощью оригинальной программы MSite 21.04.22 с использованием алгоритма, описанного в работе [5], было построено по 27 пространств для множественного молекулярного докинга, охватывающих весь объём данного белка-мишени.

### *Множественный ансамблевый молекулярный докинг известных анксиолитических соединений*

Ансамблевый докинг выполняли по методике [13] с помощью программы AutoDock Vina 1.1.1 [14], с пятикратным повтором для каждого лиганда,

каждый раз в десяти биологически активных конформациях, с вычислением по 50 полученным значениям пяти минимальных энергий связывания  $\Delta E$ . В результате для каждого соединения получали спектр из  $22 \times 27 \times 5 = 22 \times 135 = 2970$  значений  $\Delta E$ , отражающий мультитаргетный множественный аффинитет данного соединения в отношении 22 релевантных белков-мишеней. Всего полная матрица мультитаргетной множественной аффинности 537 соединений обучающей выборки включала 1594890 значений энергий докинга.

*Корреляционная свёртка спектров энергий множественного докинга*

Для одной биомишени полученные в результате множественного докинга 135 значений минимальных энергий связывания  $\Delta E$  могут быть представлены в виде полносвязной нейронной сети с симметричной матрицей связей. Эти 135 параметров выступают как нейроны и являются взаимосвязанными, поскольку рассчитаны для одного и того же белка-мишени и взаимодействие лиганда с одной из 27 непересекающихся областей докинга приведёт к конформационным изменениям белка и будет влиять на взаимодействие этого лиганда с другой областью докинга. Зависимости между ними в полносвязной нейросети полагаются линейными и могут быть представлены значениями коэффициентов парных корреляций. В корреляционной полносвязной нейросети эти коэффициенты являются весами синапсов (межнейронных связей).

С учётом этих условий, для каждой биомишени параметр свёртки спектра энергий множественного докинга представляет собой энергию  $W$  нейронной полносвязной корреляционной сети

$$W_l = \frac{1}{2} \sum_{\substack{i,j=1 \\ i \neq j}}^M R_{ij} \times \Delta E_{il} \times \Delta E_{jl}, \quad l=1 \dots N \quad (1),$$

где  $R_{ij}$  — коэффициент корреляции Пирсона между энергиями  $\Delta E_i$  и  $\Delta E_j$ ,  $i \neq j$ ;

$\Delta E_{il}$  — значение энергии  $i$  для соединения  $l$ ,  $l=1 \dots N$ ;

$\Delta E_{jl}$  — значение энергии  $j$  для соединения  $l$ ,  $l=1 \dots N$ ;

$M$  — число сворачиваемых значений энергии, равно 135;

$N$  — число соединений (лигандов).

Необходимо пояснить следующее. Значения минимальных энергий связывания практически всегда  $\Delta E \ll -1$ , а коэффициент корреляции Пирсона  $R$  изменяется в пределах от -1 до +1. Следовательно, вычисленное по формуле (1) значение  $W$  будет всегда положительным.

Таким образом, в результате свёртки мультитаргетный множественный аффинитет каждого соединения в отношении релевантных биомишеней был представлен 22 свёрточными переменными.

*Обучение нейронных сетей*

Необходимая для нейросетевого моделирования классификационная обучающая выборка включала

в себя 31 показатель: 1) шифры соединений; 2) градированные значения анксиолитической активности соединений; 3) 22 свёрточных переменных; 4) семь переменных сэмпинга. Уровень анксиолитической активности соединений обозначался метками  $hm$  (высокая или умеренная) и  $nhm$  (низкая или отсутствие активности). Свёрточные переменные были рассчитаны по формуле (1) и отражали мультитаргетный множественный аффинитет каждого соединения в отношении 22 релевантных биомишеней. Семь переменных сэмпинга задавали варианты формирования обучающей, тестовой и валидационной подвыборок в соотношении 5:1:1 и использовались для построения различных вариантов нейронных сетей.

Обучение нейросетей выполняли с помощью программы Statistica 7 [15] по схеме, изложенной в работе [13]. В соответствии с теоремой Колмогорова [16], с помощью двухслойной искусственной нейронной сети может быть аппроксимирована зависимость любой сложности. При этом желательно обеспечить свёртку сигналов входных нейронов в меньшее число промежуточных образов. Поэтому в качестве архитектуры нейронной сети был выбран двухслойный перцептрон MLP  $k-m-2$  с узким горлом. Здесь  $k$  — число входных нейронов, в данном случае 22;  $m$  — число скрытых нейронов, устанавливается программой от 3 до 20, поскольку  $2 < m < k$ . При построении классификационных нейросетей в качестве функции ошибки целесообразно использовать перекрёстную энтропию [17]. В этом случае активационной функцией для выходных нейронов является многомерная логистическая функция, а для скрытых нейронов активационными функциями могут быть линейная, логистическая, гиперболический тангенс или экспоненциальная.

Обучение нейросетей проводили с использованием алгоритма обратного распространения ошибок, перебором четырёх различных активационных функций для скрытого слоя нейронов и с использованием семи вариантов сэмпинга.

Обучение выполняли в два этапа. На первом этапе для каждого варианта сэмпинга обучали по 4000 сетей, с автоматическим отбором 200 лучших сетей. Из этих сетей по совокупности характеристик точности обучения, тестирования и валидации вручную отбирали 5 лучших. Из них по показателям точности прогноза на объединённой выборке и по результатам ROC-анализа вручную отбирали одну лучшую. На втором этапе эту нейросеть подвергали дообучению в ручном режиме с фиксированной архитектурой, с построением 200 нейросетей, из которых программой отбирались 20 лучших. Из них по показателям точности прогноза на объединённой выборке и по результатам ROC-анализа вручную отбирали одну самую лучшую.

Всего на 7 вариантах сэмпинга было обучено около 30 тыс. нейросетей и найдено 14 лучших по точности нейросетей. Из них по показателям точности прогноза на объединённой выборке и по результатам ROC-анализа была выбрана одна наилучшая.

В качестве интегральных показателей точности для всех лучших нейросетей на объединённой выборке были рассчитаны общая точность прогноза Асс, чувствительность Sens (точность прогноза активных соединений), специфичность Spec (точность прогноза неактивных соединений) и по данным ROC-анализа рассчитана площадь под кривой AUC.

Все расчёты были выполнены на суперкомпьютерных вычислительных комплексах гибридной архитектуры общей пиковой производительностью ~40 Тфлопс.

## РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

В рамках предварительной подготовки данных были построены 537 оптимизированных 3D-моделей соединений из оригинальной верифицированной базы данных [7] по структуре и анксиолитической активности известных экспериментально изученных веществ.

По описанной выше методике были найдены 22 биомешены, наиболее вероятно релевантные анксиолитической активности химических соединений, и выявлены 22 валидных 3D-модели этих белков (табл. 1).

Обращает на себя внимание тот факт, что в число выявленных вычислительным способом релевантных биомешены вошли такие “классические” для проявления анксиолитической активности белки, как ГАМК<sub>A</sub>-рецептор, серотониновые 5-HT<sub>1A</sub>- и 5-HT<sub>2A</sub>-рецепторы,  $\alpha_{2A}$ - и  $\alpha_{2B}$ -адренорецепторы, глутаматный NMDA-рецептор. Это свидетельствует об адекватности и валидности используемой методики выявления релевантных биомешены.

Таблица 1. Релевантные анксиолитической активности химических соединений биомешены и их валидированные 3D-модели

Шифр <sup>1</sup>	Название <sup>2</sup>	PDB код 3D-модели
ADRA1A	Alpha-1A adrenergic receptor	3p0g
ADRA1B	Alpha-1B adrenergic receptor	4amj
ADRA2A	Alpha-2A adrenergic receptor	6kuy
ADRA2B	Alpha-2B adrenergic receptor	3pbl
AGTR1	Type-1 angiotensin II receptor	6os1
CA2	Carbonic anhydrase 2	2weg
CA4	Carbonic anhydrase 4	5jn9
CNR1	Cannabinoid receptor 1	7v3z
GABAR	Gamma-aminobutyric acid A receptor (2 $\alpha$ 1/2 $\beta$ 2/ $\gamma$ 2)	6x3x
HTR1A	5-hydroxytryptamine receptor 1A	7e2x
HTR1B	5-hydroxytryptamine receptor 1B	4iar
HTR1D	5-hydroxytryptamine receptor 1D	5d5a
HTR2A	5-hydroxytryptamine receptor 2A	4amj
HTR2B	5-hydroxytryptamine receptor 2B	6j20
HTR2C	5-hydroxytryptamine receptor 2C	4amj
HTR4	5-hydroxytryptamine receptor 4	2rh1
HTR7	5-hydroxytryptamine receptor 5	7e2z
MTNR1A	Melatonin receptor type 1A	7vgz
MTNR1B	Melatonin receptor type 1B	7vh0
NMDAR	N-methyl-D-aspartate receptor (2GRIN1/GRIN2A/GRIN2B)	6irh
SCN11A	Sodium channel protein type 11 subunit alpha	6a90
SLC18A2	Synaptic vesicular amine transporter	3o7q

Примечание: 1 – общепринятые сокращённые обозначения белков по данным UniProt (<https://www.uniprot.org/uniprot/>); 2 – номенклатурное название, рекомендуемое UniProt (<https://www.uniprot.org/uniprot/>).

Обобщённые результаты корреляционной свёртки спектров энергий множественного ансамблевого молекулярного докинга 537 известных анксиолитических соединений в валидные 3D-модели 22 релевантных биомешены приведены в таблице 2.

Как видно из таблицы 2, для всех биомешены средние значения свёрточных переменных в классе соединений с выраженной активностью с очень высокой статистической достоверностью превышают средние значения свёрточных переменных в классе соединений с низкой активностью. Это доказывает, что корреляционная свёртка спектров энергий множественного ансамблевого молекулярного докинга в виде вычисленной по формуле (1) энергии полносвязной нейронной сети является весьма достоверной метрикой аффинности лигандов к биомешены.

В таблице 3 приведена архитектура и показатели точности наилучшей нейронной сети, полученной для выраженной анксиолитической активности при итеративном обучении в два этапа на семи вариантах сэмплинга.

По совокупности показателей точности найденная нейросеть статистически очень высоко достоверна: для пяти показателей из семи её значимость  $p < 1 \times 10^{-15}$ , а для двух  $p = 0,000170$  и  $p = 0,0165$  по биномиальному критерию [18].

Таким образом, на примере анксиолитической активности показана адекватность и высокая валидность следующих новых разработанных методик: 1) выявления биомешены, потенциально релевантных изучаемой активности; 2) множественного молекулярного докинга, не требующего нахождения в биомешены специфических сайтов связывания;

## НЕЙРОСЕТЕВАЯ МОДЕЛЬ АНКСИОЛИТИЧЕСКОЙ АКТИВНОСТИ

Таблица 2. Результаты корреляционной свёртки спектров энергий множественного ансамблевого молекулярного докинга известных анксиолитических соединений в валидные 3D-модели релевантных белков-мишеней

Шифр <sup>1</sup>	Среднее значение свёртки, W <sup>2</sup>			p <sup>6</sup>
	W <sub>hm</sub> <sup>3</sup>	W <sub>nhm</sub> <sup>4</sup>	W <sub>hm</sub> – W <sub>nhm</sub> <sup>5</sup>	
ADRA1A	12848	11849	999	3×10 <sup>-6</sup>
ADRA1B	14408	13036	1371	<5×10 <sup>-7</sup>
ADRA2A	12903	11553	1350	<5×10 <sup>-7</sup>
ADRA2B	14366	13098	1268	<5×10 <sup>-7</sup>
AGTR1	13810	12458	1352	<5×10 <sup>-7</sup>
CA2	11453	10552	901	<5×10 <sup>-7</sup>
CA4	10037	9264	773	<5×10 <sup>-7</sup>
CNR1	10225	9363	862	1×10 <sup>-6</sup>
GABAR	15664	14342	1322	<5×10 <sup>-7</sup>
HTR1A	16000	14657	1344	<5×10 <sup>-7</sup>
HTR1B	12182	11125	1057	<5×10 <sup>-7</sup>
HTR1D	13928	12621	1308	<5×10 <sup>-7</sup>
HTR2A	15544	14197	1346	1×10 <sup>-6</sup>
HTR2B	10311	9283	1028	<5×10 <sup>-7</sup>
HTR2C	14822	13558	1265	1×10 <sup>-6</sup>
HTR4	16179	14581	1598	<5×10 <sup>-7</sup>
HTR7	11322	10450	872	1×10 <sup>-6</sup>
MTNR1A	9332	8626	706	1×10 <sup>-6</sup>
MTNR1B	14660	13442	1218	<5×10 <sup>-7</sup>
NMDAR	25090	22544	2546	<5×10 <sup>-7</sup>
SCN11A	17478	15721	1756	<5×10 <sup>-7</sup>
SLC18A2	13293	12300	993	1×10 <sup>-6</sup>

Примечание: 1 – обозначения биомшеней соответствуют таблице 1; 2 – среднее вычисленных по формуле (1) энергий полносвязной корреляционной нейросети; 3 – для класса соединений с выраженной активностью; 4 – для класса соединений с низкой активностью; 5 – разность средних значений W для классов соединений с выраженной и низкой активностями; 6 – достоверность различий средних значений W по критерию Манна-Уитни [18].

3) корреляционной свёртки спектров энергий множественного докинга, основанной на вычислении энергий полносвязной симметричной нейронной сети; 4) обобщения с использованием искусственных нейронных сетей для набора релевантных биомшеней совокупности полученных свёрточных переменных, с формированием высокоточной модели, способной эффективно прогнозировать системные виды фармакологической активности.

### ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Методом искусственных нейронных сетей построена высокоточная мультитаргетная модель зависимости анксиолитической активности

Таблица 3. Архитектура и показатели точности мультитаргетной нейросетевой модели зависимости анксиолитической активности химических соединений от параметров корреляционной свёртки спектров энергий их множественного докинга

Общие характеристики лучшей нейросети		
Номер <sup>1</sup>	Архитектура <sup>2</sup>	
4/2/66	MLP 22-19-2 (Tanh, Softmax)	
Точность при обучении <sup>3</sup>		
Тип данных	F <sub>0</sub> , % <sup>4</sup>	p <sup>5</sup>
Обучающая подвыборка	98,4	<1×10 <sup>-15</sup>
Тестовая подвыборка	79,2	1,70×10 <sup>-4</sup>
Валидационная подвыборка	67,5	1,65×10 <sup>-2</sup>
Точность при тестировании на объединённой выборке		
Показатель	F, %	p <sup>5</sup>
Общая точность Acc, % <sup>6</sup>	91,2	<1×10 <sup>-15</sup>
Чувствительность Sens, % <sup>7</sup>	91,3	<1×10 <sup>-15</sup>
Специфичность Spec, % <sup>8</sup>	91,2	<1×10 <sup>-15</sup>
Площадь под ROC-кривой AUC, % <sup>9</sup>	94,4	<1×10 <sup>-15</sup>

Примечание: 1 – указаны номер этапа обучения, номер сэмпинга и номер нейросети; 2 – многослойный перцептрон, 22 входных, 19 скрытых и 2 выходных нейрона, функции активации скрытого и выходного слоя гиперболический тангенс и многомерная логистическая функция; 3 – получены в двухэтапном итеративном процессе построения нейросети; 4 – общая точность Acc; 5 – достоверность показателей точности по биномиальному критерию [18]; 6 – точность прогноза для всех соединений выборки; 7 – точность прогноза для соединений с выраженной активностью; 8 – точность прогноза для мало активных соединений; 9 – для всей выборки.

химических соединений от их интегральной аффинности к 22 релевантным белкам-мишеням, основанная на корреляционной свёртке спектров энергий множественного докинга.

Точность прогноза анксиолитической активности с применением построенной классификационной мультитаргетной свёрточной нейросетевой модели является весьма высокой и в большинстве тестов превышает 90%. Модель статистически очень достоверна, в большинстве тестов её значимость составляет  $p < 1 \times 10^{-15}$ .

Таким образом, на примере анксиолитической активности разработана новая методология построения классификационных моделей для прогноза системных видов мультитаргетной активности химических соединений, основанная на использовании технологии искусственных нейронных сетей и на корреляционной свёртке спектров энергий множественного докинга.

Построенная мультитаргетная нейросетевая модель зависимости анксиолитической активности химических соединений от параметров корреляционной свёртки спектров энергий множественного докинга используется в поиске

*in silico* новых высокоактивных соединений различных химических классов, найдены перспективные вещества.

Созданная мультитаргетная свёрточная нейросетевая методология может быть использована для поиска высокоактивных соединений с другими видами фармакологической активности, имеющими общесистемный характер, такими, как гипогликемическая, противовоспалительная, иные виды психотропной активности.

## ФИНАНСИРОВАНИЕ

Работа выполнена в рамках государственного задания Министерства здравоохранения Российской Федерации № 23022400009-9 “Разработка методологии компьютерного поиска мультитаргетных фармакологически активных соединений на основе множественного докинга и технологии свёрточных нейронных сетей различной архитектуры”.

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая статья не содержит описания проведённых авторами исследований с участием людей или использованием животных в качестве объектов.

## КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

## ЛИТЕРАТУРА

- World Health Organization. Mental Health, Brain Health and Substance Use. Retrieved July 29, 2024, from: <https://www.who.int/teams/mental-health-and-substance-use/overview>
- Здравоохранение в России. 2023: Статистический сборник (2023), Росстат, Москва, 179 с. [Zdravookhraneniye v Rossii. 2023: Statisticheskiy Sbornik (2023), Rosstat, Moskva, 179 p.]
- Sarkar C., Das B., Rawat V.S., Wahlang J.B., Nongpiur A., Tiewsoh I., Lyngdoh N.M., Das D., Bidarolli M., Sony H.T. (2023) Artificial intelligence and machine learning technology driven modern drug discovery and development. *Int. J. Mol. Sci.*, **24**(3), 2026. DOI: 10.3390/ijms24032026
- Vora L.K., Gholap A.D., Jetha K., Thakur R.R.S., Solanki H.K., Chavda V.P. (2023) Artificial intelligence in pharmaceutical technology and drug delivery design. *Pharmaceutics*, **15**(7), 1916. DOI: 10.3390/pharmaceutics15071916
- Васильев П.М., Кочетков А.Н., Спасов А.А., Перфильев М.А. (2021) Спектр энергий множественного докинга как многомерная метрика аффинности химических соединений к фармакологически релевантным биомишеням. *Волгоградский научно-медицинский журнал*, **3**, 57–61. [Vassiliev P.M., Kochetkov A.N., Spasov A.A., Perfilev M.A. (2021) The energy spectrum of multiple docking as a multi-dimensional metric of the affinity of chemical compounds to pharmacologically relevant bio-targets. *Volgogradskiy Nauchno-Meditsinskiy Zhurnal*, **3**, 57–61.]
- Васильев П.М., Кочетков А.Н., Перфильев М.А. (2022) Нейросетевое моделирование зависимости ГАМК<sub>A</sub>-агонистической активности химических соединений от спектра энергий множественного докинга. *Вестник ВолгГМУ*, **19**(4), 88–93. [Vassiliev P.M., Kochetkov A.N., Perfilev M.A. (2022) Neural network modeling of the dependence of GABA<sub>A</sub>-agonistic activity of chemical compounds on the energy spectrum of multiple docking. *Vestnik VolgGMU*, **19**(4), 88–93.] DOI: 10.19163/1994-9480-2022-19-4-88-93
- Васильев П.М., Мальцев Д.В., Перфильев М.А., Спасов А.А., Скрипка М.О., Кочетков А.Н. (2022) Свидетельство о государственной регистрации базы данных № 2022621744. Соединения с анксиолитической активностью (15.07.2022), Офиц. бюлл. Программы для ЭВМ. БД. ТИМС, **7**, RU 2022621744. [Vassiliev P.M., Maltsev D.V., Perfilev M.A., Spasov A.A., Skripka M.O., Kochetkov A.N. (2022) Soyedineniya s anksioliticheskoy aktivnost'yu (Compounds with anxiolytic activity), Certificate of State Registration of the Database no. 2022621744. (15.07.2022), Offits. Bull. Programmy dlya EVM, baz dannykh i topologiy integral'nykh mikroskhem ETIN (Programs for Computers, Databases and Topologies of Integrated Circuits), **7**, RU 2022621744.]
- ChemAxon: Marvin (2024), Retrieved July 29, 2024, from: <https://chemaxon.com/marvin>
- MOPAC (2024), Retrieved July 29, 2024, from: <http://openmopac.net/>
- Ochoa D., Hercules A., Carmona M., Suveges D., Baker J., Malangone C., Lopez I., Miranda A., Cruz-Castillo C., Fumis L., Bernal-Llinares M., Tsukanov K., Cornu H., Tsirigos K., Razuvayevskaya O., Buniello A., Schwartzentruber J., Karim M., Ariano B., Martinez Osorio R.E., Ferrer J., Ge X., Machlitt-Northen S., Gonzalez-Uriarte A., Saha S., Tirunagari S., Mehta C., Roldán-Romero J.M., Horswell S., Young S., Ghoussaini M., Hulcoop D.G., Dunham I., McDonagh E.M. (2023) The next-generation Open Targets Platform: reimaged, redesigned, rebuilt. *Nucleic Acids Res.*, **51**(D1), D1353–D1359. DOI: 10.1093/nar/gkac1046
- Vasilyev P.M., Luzina O.A., Babkov D.A., Appazova D.T., Salakhutdinov N.F., Spasov A.A. (2019) Studying dependences between the chemotype structure of some natural compounds and the spectrum of their targeted activities correlated with the hypoglycemic effect. *J. Struct. Chem.*, **60**(11), 1827–1832. DOI: 10.1134/S0022476619110179
- Vassiliev P.M., Spasov A.A., Kosolapov V.A., Kucheryavenko A.F., Gurova N.A., Anisimova V.A. (2014) Consensus drug design using it microcosm. In: *Application of Computational Techniques in Pharmacy and Medicine* (Gorb L., Kuz'min V., Muratov E., eds.). Springer Science + Business Media, Dordrecht (Netherlands), pp. 369–431.
- Васильев П.М., Спасов А.А., Яналиева Л.Р., Кочетков А.Н., Ворфоломеева В.В., Ключков В.Г., Анназова Д.Т. (2019) Нейросетевое моделирование мультитаргетной RAGE-ингибирующей активности. *Биомедицинская химия*, **65**(2), 91–98. [Vassiliev P.M., Spasov A.A., Yanaliyeva L.R., Kochetkov A.N., Vorfolomeyeva V.V., Klockhov V.G., Appazova D.T. (2019) Neural network modeling of multitarget RAGE inhibitory activity. *Biomeditsinskaya Khimiya*, **65**(2), 91–98.] DOI: 10.1134/S1990750819030107
- Trott O., Olson A.J. (2010) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**(2), 455–461. DOI: 10.1002/jcc.21334

15. Hilbe J.M. (2007) Statistica 7. An overview. The American Statistician, **61**(1), 91–94.
16. Колмогоров А.Н. (1957) О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. Докл. АН СССР, **114**(5), 953–956. [Kolmogorov A.N. (1957) On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. Proceedings of USSR Academy of Sciences, **114**(5), 953–956.]
17. Dubin U. (2013) Cross-Entropy Method: Theory with Applications. LAMBERT Academic Publishing, Chisinau (Moldova), 148 p.
18. Глотов Н.В., Животовский Л.А., Хованов Н.В., Хромов-Борисов Н.Н. (1982) Биометрия, Изд-во Ленингр. ун-та, Л., 264 с. [Glotov N.V., Zhivotovskij L.A., Hovanov N.V., Hromov-Borisov N.N. (1982) Biometriya, Izd-vo Leningr. Un-ta, L., 264 p.]

Поступила в редакцию: 01. 10. 2024.  
После доработки: 25. 10. 2024.  
Принята к печати: 27. 10. 2024.

## MULTI-TARGET NEURAL NETWORK MODEL OF ANXIOLYTIC ACTIVITY OF CHEMICAL COMPOUNDS USING CORRELATION CONVOLUTION OF MULTIPLE DOCKING ENERGY SPECTRA

*P.M. Vassiliev\*, M.A. Perfilev, A.V. Golubeva, A.N. Kochetkov, D.V. Maltsev*

Volgograd State Medical University,  
1 Pavshikh Bortsov sq., Volgograd, 400131 Russia; \*e-mail: pvassiliev@mail.ru

Anxiety disorders are one of the most common mental health pathologies in the world. They require search and development of novel effective pharmacologically active substances. Thus, the development of new approaches to the search for anxiolytic substances by artificial intelligence methods is an important area of modern bioinformatics and pharmacology. In this work, a multi-target model of the dependence of the anxiolytic activity of chemical compounds on their integral affinity to relevant target proteins based on the correlation convolution of multiple docking energy spectra has been constructed using the method of artificial neural networks. The training set of the structure and activity of 537 known anxiolytic substances was formed on the basis of the previously created database, and optimized 3D models of these compounds were built. 22 biotargets presumably relevant to anxiolytic activity were identified and their valid 3D models were found. For each biotarget, 27 multiple docking spaces have been formed throughout its entire volume. Multiple ensemble molecular docking of 537 known anxiolytic compounds into all spaces of relevant target proteins has been performed. The correlation convolution of the calculated energy spectra of multiple docking was carried out. Using seven training options based on artificial multilayer perceptron neural networks, the multi-target model of depending anxiolytic activity chemical compounds on 22 parameters of the correlation convolution of the multiple docking spectra energy was constructed. The predictive ability of the created model was characterized  $\text{Acc} = 91.2\%$  and  $\text{AUC}_{\text{ROC}} = 94.4\%$ , with statistical significance of  $p < 1 \times 10^{-15}$ . The found model is currently used in the search for new substances with high anxiolytic activity.

*The whole English version is available at <http://pbmc.ibmc.msk.ru>.*

**Key words:** anxiolytic activity; multi-target affinity; multiple molecular docking; docking energy spectra; correlation convolution; artificial neural networks

**Funding.** This study was supported in the framework of the state assignment of the Ministry of Health of the Russian Federation No. 23022400009-9 “Development of a methodology for computer searching multi-target pharmacologically active compounds based on multiple docking and technology of the convolutional neural networks of various architectures”.

Received: 01.10.2024; revised: 25.10.2024; accepted: 27.10.2024.