

Читать
онлайн
Read
onlineГусева Е.А.^{1,2}, Николаева Н.И.², Филин А.С.², Савостикова О.Н.¹

Сравнительная оценка математических моделей прогнозирования острой токсичности химических веществ

¹ФГБУ «Центр стратегического планирования и управления медико-биологическими рисками здоровью» Федерального медико-биологического агентства, 119121, Москва, Россия;

²ФГАОУ ВО «Первый Московский государственный медицинский университет имени И.М. Сеченова» Министерства здравоохранения Российской Федерации (Сеченовский Университет)», 199911, Москва, Россия

Введение. Оценке острой токсичности химических соединений при пероральном поступлении уделяется значительное внимание в связи с различной скоростью всасывания веществ у разных видов животных и разными условиями проведения эксперимента. Учитывая темпы развития химической промышленности, перед исследователями встает вопрос об ускорении изучения свойств веществ и заполнении пробелов в данных. Поэтому прогнозирование на количественном уровне токсических свойств веществ с помощью математических моделей на основе структуры или структурных свойств соединений — QSAR-моделирование — является одним из перспективных направлений.

Цель исследования — создание и сравнение полученных математических моделей для прогнозирования острой токсичности химических веществ различных классов.

Материалы и методы. В исследование было включено четыре класса пестицидов (хлорорганические соединения (ХОС), азолы, карбаматы, фосфорорганические соединения (ФОС)) в количестве 100 соединений с дескрипторами, рассчитанными программным обеспечением PaDEL-Descriptors ver. 2.21. В программе WEKA были построены модели регрессии, подвергнутые процедуре внутренней валидации. Для оценки качества регрессионных моделей были использованы статистические параметры: среднеквадратичная ошибка (RMSE) и коэффициент детерминации (r^2).

Результаты. Для прогнозирования острой пероральной токсичности ХОС и ФОС оптимально использование модели, в которой происходит комбинирование нейронных сетей и метода опорных векторов, для карбаматов — ансамблевой модели, включающей в себя линейную регрессию и метод опорных векторов. Для веществ из группы азолов не удалось создать модели, которая бы соответствовала необходимым требованиям: $r^2 > 0,6$ для тренировочного набора и $r^2 > 0,5$ при проведении кросс-валидации.

Ограничения исследования. Исследование ограничено количеством исследуемых соединений, классами химических соединений, областью распространения полученных в ходе моделирования результатов.

Заключение. В данном исследовании ансамблевые методы моделирования продемонстрировали наилучшие результаты при прогнозировании острой пероральной токсичности для ХОС, карбаматов, ФОС.

Ключевые слова: QSAR; острая токсичность; математические модели; прогнозирование

Соблюдение этических стандартов. Исследование не требует представления заключения комитета по биомедицинской этике или иных документов.

Для цитирования: Гусева Е.А., Николаева Н.И., Филин А.С., Савостикова О.Н. Сравнительная оценка математических моделей прогнозирования острой токсичности химических веществ. *Гигиена и санитария*. 2022; 101(7): 816–823. <https://doi.org/10.47470/0016-9900-2022-101-7-816-823> <https://www.elibrary.ru/trwbtp>

Для корреспонденции: Гусева Екатерина Андреевна, специалист отдела физико-химических исследований и экотоксикологии, ФГБУ «ЦСП» ФМБА, 119121, Москва, Россия; аспирант, ассистент кафедры экологии человека и гигиены окружающей среды Института общественного здоровья им. Ф.Ф. Эрисмана, ФГАОУ ВО «Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет)», 199911, Москва. E-mail: guseva_e_a@staff.sechenov.ru

Участие авторов: Гусева Е.А. — концепция и дизайн исследования, сбор материала и обработка данных, статистическая обработка, написание текста; Николаева Н.И. — написание текста, редактирование; Филин А.С. — редактирование; Савостикова О.Н. — редактирование. Все соавторы — утверждение окончательного варианта статьи, ответственность за целостность всех частей статьи.

Конфликт интересов. Авторы декларируют отсутствие явных и потенциальных конфликтов интересов в связи с публикацией данной статьи.

Финансирование. Исследование не имело финансовой поддержки.

Поступила: 18.03.2022 / Принята к печати: 08.06.2022 / Опубликована: 31.07.2022

Ekaterina A. Guseva^{1,2}, Natalia I. Nikolayeva², Andrey S. Filin², Olga N. Savostikova¹

Comparative evaluation of mathematical models for predicting acute toxicity of chemicals

¹Centre for Strategic Planning of FMBA of Russia, Moscow, 119121, Russian Federation;

²I.M. Sechenov First Moscow State Medical University of the Ministry of Health of the Russian Federation (Sechenov University), Moscow, 199911, Russian Federation

Introduction. Considerable attention is paid to the assessment of acute toxicity of chemical compounds during oral administration due to the different rates of absorption of substances in different animal species and various experimental conditions. Given the pace of development of the chemical industry, researchers are faced with the question of accelerating the study of the properties of substances and filling data gaps. Therefore, quantitative prediction of the toxic properties of substances using mathematical models based on the structure or structural properties of compounds — quantitative structure — activity relationship (QSAR) modeling — is one of the promising areas.

The purpose of this study is to create and compare the performance of the obtained mathematical models for predicting the acute toxicity of various classes of chemicals.

Materials and methods. The study included four classes of pesticides (organochlorine compounds (OCs), azoles, carbamates, organophosphorus compounds (OPs)) in the amount of 100 compounds with descriptors calculated by PaDEL-Descriptors software ver. 2.21. Regression models were constructed in the WEKA software, subjected to an internal validation procedure. Statistical parameters such as the mean square error (RMSE) and the coefficient of determination (r^2) were used to assess the quality of regression models.

Results. To predict acute oral toxicity of OCs and OPs, it is optimal to use a model in which neural networks and the support vector method are combined, for carbamates — an ensemble model that includes linear regression and the support vector method. For substances from theazole group, it was not possible to create a model that would meet the necessary requirements: $r^2 > 0.6$ for the training set and $r^2 > 0.5$ for cross-validation.

Limitations. The study is limited by the number of compounds studied, the class of chemical compounds, and the area of distribution of the results obtained during modeling.

Conclusion. In this study, ensemble modelling methods demonstrated the best results in predicting acute oral toxicity for OCs, carbamates, and OPs.

Keywords: QSAR; acute toxicity; mathematical models; prediction

Compliance with ethical standards. The study does not require submission of the opinion of the biomedical ethics committee or other documents.

For citation: Guseva E.A., Nikolayeva N.I., Filin A.S., Savostikova O.N. Comparative evaluation of mathematical models for predicting acute toxicity of chemicals. *Gigiena i Sanitariya (Hygiene and Sanitation, Russian journal)*. 2022; 101(7): 816–823. <https://doi.org/10.47470/0016-9900-2022-101-7-816-823> <https://elibrary.ru/trwbtp> (in Russian)

For correspondence: Ekaterina A. Guseva, Specialist of the Department of Physico-Chemical Research and Ecotoxicology, Centre for Strategic Planning of FMBA of Russia, Moscow, 119121, Russian Federation; postgraduate student, Assistant of the Department of Human Ecology and Environmental Hygiene of the Institute of Public Health named after F.F.Erisman, Sechenov First Moscow State Medical University of the Ministry of Health of Russia (Sechenov University), Moscow, 199911, Russian Federation. E-mail: guseva_e_a@staff.sechenov.ru

Information about the authors:

Guseva E.A., <https://orcid.org/0000-0001-8389-7981>

Nikolayeva N.I., <https://orcid.org/0000-0003-1226-9990>

Filin A.S., <https://orcid.org/0000-0002-9724-8410>

Savostikova O.N., <https://orcid.org/0000-0002-7032-1366>

Contribution: Guseva E.A. — the concept and design of the study, collection and processing of material, writing a text; Nikolayeva N.I. — writing a text, editing; Filin A.S. — editing; Savostikova O.N. — editing. All authors are responsible for the integrity of all parts of the manuscript and approval of the manuscript final version.

Conflict of interest. The authors declare no conflict of interest.

Acknowledgement. The study had no sponsorship.

Received: March 18, 2022 / Accepted: June 08, 2022 / Published: July 31, 2022

Введение

Оценке острой токсичности химических соединений при пероральном поступлении уделяется значительное внимание в связи с рядом особенностей, устанавливающих сложность данного процесса [1]. Во-первых, вещества имеют разную скорость всасывания и метаболизма и соответственно разную биодоступность. Данный показатель зависит не только от свойств самого вещества, но и от того, поступает ли данное вещество изолированно или вместе с водой и пищей, поступает изначально в полость рта или непосредственно через зонд в желудок, а также ряда других факторов. Таким образом, возможна регистрация многочисленных вариантов токсического действия на опытных животных. Во-вторых, токсичность соединений зависит от пола и вида экспериментального животного. В-третьих, данные, полученные в разных лабораториях, зачастую имеют отличия из-за возможных модификаций тестов [2].

Для классификации химических веществ с точки зрения их потенциальной опасности для здоровья человека используются тесты на острую пероральную токсичность на млекопитающих. В настоящее время для этих целей используются тесты, одобренные Организацией экономического сотрудничества и развития (ОЭСР), ратифицированные на территории Российской Федерации: № 420 «Острая токсичность при внутрижелудочном поступлении — метод фиксированной дозы», № 423 «Острая пероральная токсичность — метод определения класса острой токсичности», № 425 «Острая пероральная токсичность — метод “вверх и вниз”» [3]. При проведении вышеупомянутых тестов исследуемое химическое вещество вводится, как правило, женским особям крыс через желудочный зонд либо однократной дозой, либо дозу вводят частями в течение 24 ч. Ранее для оценки использовался тест ОЭСР № 401, но по этическим соображениям в 2002 г. он был отменён. Несмотря на то что количество животных, используемых в экспериментах, значительно снижается, тесты *in vivo* по-прежнему остаются энерго- и ресурсозатратными [4, 5]. По данным Химической реферативной службы CAS (подразделения Американского химического общества), по состоянию на май 2022 г. зарегистрировано около 194 млн органических и неорганических веществ, включая сплавы, координационные соединения, минералы, смеси, полимеры и соли [6], каждое из которых обладает физико-химическими характеристиками, определяющими их токсические свойства. С учётом значительного расширения использования химических веществ перед исследователями встаёт вопрос об ускорении изучения свойств веществ и за-

полнении пробелов в данных. Поэтому прогнозирование на количественном уровне токсических свойств веществ с помощью математических моделей на основе структуры или структурных свойств соединений (QSAR-моделирование, англ. *quantitative structure-activity relationship*) является одним из перспективных направлений [7–9]. QSAR-модель должна иметь определённую конечную точку, однозначную стратегию построения модели с учётом характера выбранных данных, определённую область применения, соответствующие показатели пригодности, надёжности и прогнозируемости и, наконец, предлагать возможную механистическую интерпретацию разработанных моделей [10].

Для построения QSAR-модели необходимы 3 компонента:

- 1) набор данных, обеспечивающий экспериментальные измерения биологической активности или свойства для группы уже испытанных химических веществ (обучающий набор);
- 2) данные о молекулярной структуре и/или свойствах моделируемых соединений (дескрипторы);
- 3) статистические методы для поиска и проверки взаимосвязи между этими двумя наборами [11–13].

Цель исследования — создание и сравнение полученных математических моделей для прогнозирования острой токсичности различных классов химических веществ.

Материалы и методы

Все тесты по исследованию острой пероральной токсичности направлены на установление DL_{50} (мг/кг) при внутрижелудочном поступлении, поэтому данный параметр стал подходящей конечной точкой для целей моделирования.

В связи с широкой распространённостью пестицидов и частым контактом с ними человека эти вещества являются наиболее изученными с токсиколого-гигиенической точки зрения [14]. Поэтому в работе был использован набор данных о токсичности ряда веществ, организованный на основе их химической классификации:

- 1) хлорорганические соединения (ХОС) — галопроизводные полициклических углеводородов и углеводородов алифатического ряда;
- 2) азолы — пятичленные гетероциклы, имеющие в цикле не менее двух гетероатомов, один из которых атом азота, а также би- и полициклические соединения, включающие азольный цикл;
- 3) карбаматы — сложные эфиры карбаминовой кислоты (NH_2COOH);

Таблица 1 / Table 1

Распределение дескрипторов, необходимых для моделирования
Distribution of descriptors required for modelling

Модель Model	Выбранные дескрипторы Selected descriptors
ХОС Chloroorganics	ATS6m, ATS8p, ATSC8m, AATSC7i, AATSC8i, MATS6v, MATS3i, MATS7i, MATS8i, nsCH3, SHCsats, MIC0, MDEC-13, n5Ring, n5HeteroRing, nF8HeteroRing, nF9HeteroRing, nT5HeteroRing
Азолы Azoles	ALogP, ATSC4m, AYSC7s, AATSC3m, MATS7i, MATS2s, MATS4s, GATS3m, VR2_Dzp, nHBint4, minHssNH, minssNH, maxHssNH, MIC1, nF10Ring, nT10Ring, n6HeteroRing, nF10HeteroRing
Карбаматы Carbamates	ATSC8m, ATSC7e, ATSC3p, AATSC2e, MATS6i, MATS7i, GATS6c, CIC2, MDEC-14, MDEC-22, JGI6
ФОС Organophosphorous compounds	ATS6s, ATSC6i, MATS5m, GATS3c, GATS7v, MDEC-13, LipinskiFailures, JGI7

4) фосфорорганические соединения (ФОС) – органические соединения пятивалентного фосфора, содержащие связь «водород – фосфор».

Экспериментальные значения острой токсичности при пероральном введении крысам, выраженные в DL_{50} (мг/кг), были получены из общедоступных баз данных: Базы данных свойств пестицидов (Pesticide Properties Data Base – PPDB) [15], Регистра токсических эффектов химических соединений (Registry of Toxic Effects of Chemical Substances – RTECS) [16], Базы данных ОЭСР – eChemPortal. Из набора были исключены соединения, имеющие несколько значений DL_{50} , если величина наибольшего значения в три и более раз превышала наименьшее значение. Для оставшихся веществ, имеющих несколько значений DL_{50} , было вычислено среднее геометрическое. Перед моделированием все значения были преобразованы в моль/кг и выражены как $\lg 1/LD_{50}$ в соответствии со стандартными методами QSAR [17].

В выборку не вошли неорганические соединения, смеси, металлы, полимеры, соли и химические вещества, содержащие элементы, отличные от С, Н, О, N, F, Cl, Br, I, S, P, Si, As, из-за невозможности вычисления молекулярных дескрипторов для этих соединений.

Расчёт двумерных химических дескрипторов выбранных соединений проводили с использованием компьютерной программы PaDEL-Descriptors ver. 2.21 (Yap Chun Wei, Pharmaceutical Data Exploration Laboratory) [18]. Для интеллектуального анализа данных использовали среду Waikato

для анализа знаний (WEKA) ver. 3.9.6 [19]. Были построены математические модели регрессии, которые были подвергнуты процедуре внутренней 10-кратной кросс-валидации. Для оценки качества регрессионных моделей были использованы такие статистические параметры, как среднеквадратичная ошибка (RMSE), описывающая среднюю меру ошибки в прогнозировании зависимой переменной, и коэффициент детерминации (r^2), представляющий собой процент изменчивости, который может быть объяснён моделью [20, 21].

Результаты

Для уменьшения количества избыточных и бесполезных переменных из пула дескрипторов были исключены переменные с постоянными значениями или с хотя бы одним отсутствующим значением. С помощью функции CfsSubsetEval.BestFirst был выбран ряд релевантных дескрипторов для каждого набора данных на основе их значительной корреляции с конечной точкой. Так, для ХОС было использовано 18 дескрипторов, для азолов – 18, для карбаматов – 11, для ФОС – 8 (табл. 1).

Были построены математические модели на основе таких методов, как линейная регрессия (linear regression – LR), k-ближайших соседей (k-nearest neighbor – k-NN), дерево принятия решений (random tree – RT), метод опорных векторов (support vector machine – SVM), нейронные сети (NN – neural networks). Полученные результаты представлены в табл. 2.

Таблица 2 / Table 2

Экспериментальные и прогнозируемые значения острой пероральной токсичности различных групп пестицидов, полученные в результате математического моделирования

Experimental and predicted values of acute oral toxicity of various groups of pesticides obtained as a result of mathematical modelling

Вещество Compound name	$\lg (1 / DL_{50})$, экспериментальное $\lg (1 / DL_{50})$, experimental	$\lg (1 / DL_{50})$, прогнозируемое $\lg (1 / DL_{50})$, predicted						Предел ошибки Margin of error					
		ЛР LR	k-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM	ЛР LR	k-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM
<i>Хлороорганические соединения (ХОС) / Organochlorine compounds (OCs)</i>													
Альдрин / Aldrin	3.913	4.58	4.361	4.29	4.442	4.161	4.302	0.667	0.448	0.377	0.529	0.248	0.389
Альфа-ГХЦГ / Alpha-HCG	3.216	3.395	3.42	3.281	3.346	3.11	3.228	0.179	0.204	0.065	0.13	-0.106	0.012
Гептахлор / Heptachlor	3.535	3.282	3.415	3.537	3.534	3.398	3.466	-0.253	-0.12	0.002	-0.001	-0.137	-0.069
ДДД / DDD	3.452	3.601	3.382	3.452	3.453	3.335	3.394	0.149	-0.07	-0	0.001	-0.117	-0.058
ДДТ / DDT	3.514	3.582	3.325	3.514	3.415	3.26	3.337	0.068	-0.189	0	-0.099	-0.254	-0.177
Дикофол / Dicofol	2.724	2.549	2.933	2.724	2.723	2.536	2.629	-0.175	0.209	-0	-0.001	-0.188	-0.095

Продолжение таблицы 2 на стр. 819–821

Продолжение таблицы 2. Начало на стр. 818.

Вещество Compound name	lg (1/DL ₅₀), экспериментальное lg (1/DL ₅₀), experimental	lg (1/DL ₅₀), прогнозируемое lg (1/DL ₅₀), predicted						Предел ошибки Margin of error					
		ЛР LR	к-БС k-NN	ДПП DT	МОП SVM	НС NN	АМ EM	ЛР LR	к-БС k-NN	ДПП DT	МОП SVM	НС NN	АМ EM
<i>Хлорорганические соединения (ХОС) / Organochlorine compounds (OCs)</i>													
Дильдрин / Dieldrin	3.958	4.244	4.361	4.432	4.292	4.229	4.261	0.286	0.403	0.474	0.334	0.271	0.303
Изобензан / Isobenzane	5.035	5.035	4.09	5.035	5.031	4.814	4.922	-0	-0.945	-0	-0.004	-0.221	-0.113
Изадрин / Izadrin	4.667	4.58	4.361	4.29	4.442	4.161	4.302	-0.087	-0.306	-0.377	-0.225	-0.506	-0.365
Линдан (гамма-ГХЦГ) Lindane (gamma-HCG)	3.346	3.395	3.42	3.281	3.346	3.11	3.228	0.049	0.074	-0.065	-0	-0.236	-0.118
Метоксихлор / Methoxychlor	1.957	2.212	2.17	1.957	1.953	1.793	1.873	0.255	0.213	-0	-0.004	-0.164	-0.084
Мирекс / Mirex	3.064	3.007	3.331	3.064	3.068	2.759	2.914	-0.057	0.267	0	0.004	-0.305	-0.15
Пентахлорфенол Pentachlorophenol	3.539	3.393	3.409	3.537	3.541	3.526	3.534	-0.146	-0.13	-0.002	0.002	-0.013	-0.005
Полихлорпинен (стробан) Polychloropinene (stroban)	3.276	3.351	3.503	3.276	3.282	3.088	3.185	0.075	0.227	0	0.006	-0.188	-0.091
Токсафен / Toxaphene	3.699	3.395	3.449	3.699	3.693	3.509	3.601	-0.304	-0.25	0	-0.006	-0.19	-0.098
Хлордан / Chlordane	3.01	2.89	3.415	3.01	3.064	2.802	2.933	-0.12	0.405	-0	0.054	-0.208	-0.077
Хлорпропилат / Chloropropylate	1.83	2.067	2.17	1.83	1.834	1.66	1.747	0.237	0.34	0	0.004	-0.17	-0.083
Эндосульфан / Endosulfan	3.884	3.72	4.249	3.884	3.882	3.747	3.814	-0.164	0.365	0	-0.002	-0.137	-0.07
Эндрин / Endrin	4.905	4.244	4.361	4.432	4.292	4.229	4.261	-0.661	-0.544	-0.473	-0.613	-0.676	-0.644
<i>Азолы / Azoles</i>													
Азаконазол / Azaconazole	2.989	2.802	2.65	2.989	2.761	2.988	2.913	-0.187	-0.339	0	-0.228	-0.001	-0.076
Беномил / Benomil	1.463	1.622	1.556	1.463	1.46	1.827	1.583	0.159	0.093	0	-0.003	0.364	0.12
Бромконазол / Bromconazole	3.037	2.955	2.731	3.037	3.038	3.13	3.068	-0.082	-0.306	0	0.001	0.093	0.031
Гексаконазол / Hexaconazole	2.157	2.372	2.373	2.148	2.515	2.163	2.275	0.215	0.216	-0.009	0.358	0.006	0.118
Диниконазол / Diniconazole	2.838	2.623	2.873	2.838	2.586	2.901	2.775	-0.215	0.035	0	-0.252	0.063	-0.063
Дифеноконазол Diphenconazole	2.447	2.457	2.47	2.447	2.445	2.747	2.546	0.01	0.023	0	-0.002	0.3	0.099
Имазалил / Imazalil	3.117	2.800	2.786	3.117	2.725	3.191	3.011	-0.317	-0.331	0	-0.392	0.074	-0.106
Имибенконазол Imibenconazole	2.167	2.734	2.660	2.167	2.726	2.174	2.356	0.567	0.493	0	0.559	0.007	0.189
Ипконазол / Ipconazole	2.575	2.530	2.497	2.575	2.409	2.727	2.57	-0.045	-0.078	0	-0.166	0.152	-0.005
Карбендазим / Carbendazim	1.43	1.384	1.556	1.430	1.434	1.646	1.503	-0.046	0.126	0	0.004	0.216	0.073
Климбазол / Climbazole	2.864	2.640	2.882	2.864	2.541	2.818	2.741	-0.224	0.018	0	-0.323	-0.046	-0.123
Метконазол / Metconazole	2.739	2.448	2.350	2.739	2.283	2.544	2.522	-0.291	-0.389	0	-0.456	-0.195	-0.217
Мефентрифлуконазол Mefentrifluconazole	2.299	2.255	2.295	2.299	2.287	2.580	2.389	-0.044	-0.004	0	-0.012	0.281	0.09
Оксоконазол фумарат Oxconazole fumarate	2.664	2.548	2.562	2.667	2.529	2.691	2.629	-0.116	-0.102	0.003	-0.135	0.027	-0.035
Пенконазол / Penconazole	2.139	2.412	2.367	2.148	2.345	2.611	2.368	0.273	0.228	0.009	0.206	0.472	0.229
Пробеназол / Probenazole	2.041	2.220	2.290	2.041	2.095	2.090	2.075	0.179	0.249	0	0.054	0.049	0.034
Проклораз / Prochloraz	2.469	2.382	2.482	2.469	2.471	2.547	2.496	-0.087	0.013	0	0.002	0.078	0.027

Продолжение таблицы 2 на стр. 820, 821.

Продолжение таблицы 2. Начало на стр. 818.

Вещество Compound name	lg (1/DL ₅₀), эксперимен- тальное lg (1/DL ₅₀), experimental	lg (1/DL ₅₀), прогнозируемое lg (1/DL ₅₀), predicted						Предел ошибки Margin of error					
		ЛР LR	к-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM	ЛР LR	к-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM
<i>Азолы / Azoles</i>													
Пропиконазол / Propiconazole	2.574	2.609	2.373	2.575	2.578	2.608	2.587	0.035	-0.201	0.001	0.004	0.034	0.013
Протиоконазол / Protioconazole	1.744	1.860	1.649	1.744	1.748	1.800	1.764	0.116	-0.095	0	0.004	0.056	0.02
Симеконазол / Simeconazole	2.681	2.777	2.590	2.681	2.759	2.850	2.763	0.096	-0.091	0	0.078	0.169	0.082
Тебуконазол / Tebuconazole	2.11	2.572	2.350	2.110	2.422	2.668	2.4	0.462	0.240	0	0.312	0.558	0.29
Тетраконазол / Tetraconazole	2.557	2.567	2.817	2.557	2.558	2.901	2.672	0.010	0.260	0	0.001	0.344	0.115
Тиабендазол / Thiabendazole	1.795	1.826	2.233	1.795	1.797	1.842	1.811	0.031	0.438	0	0.002	0.047	0.016
Тиофанат-метил Methyl thiophanate	1.774	1.564	1.649	1.774	1.774	1.941	1.83	-0.210	-0.125	0	-0	0.167	0.056
Триадименол / Triadimenol	2.252	2.705	2.678	2.252	2.597	2.847	2.565	0.453	0.426	0	0.345	0.595	0.313
Триадимефон / Triadimephone	2.949	2.669	2.882	2.947	2.569	2.838	2.785	-0.280	-0.067	-0.002	-0.38	-0.111	-0.164
Тритриконазол / Triticonazole	2.201	2.409	2.350	2.201	2.268	2.390	2.286	0.208	0.149	0	0.067	0.189	0.085
Трициклазол / Tricyclazole	2.847	2.711	2.549	2.847	2.845	3.338	3.01	-0.136	-0.298	0	-0.002	0.491	0.163
Униконазол / Uniconazole	2.832	2.605	2.678	2.832	2.502	2.819	2.718	-0.227	-0.154	0	-0.33	-0.013	-0.114
Фенбуконазол / Fenbuconazole	2.226	2.493	2.551	2.226	2.331	2.698	2.418	0.267	0.325	0	0.105	0.472	0.192
Флузилазол / Fluzilazole	2.67	2.896	2.576	2.667	2.843	2.890	2.8	0.226	-0.094	-0.003	0.173	0.22	0.13
Флуотримазол / Fluotrimazole	1.88	2.021	2.073	1.880	1.882	1.878	1.88	0.141	0.193	0	0.002	-0.002	-0
Флутриафол / Flutriafol	2.422	2.399	2.408	2.422	2.418	2.580	2.473	-0.023	-0.014	0	-0.004	0.158	0.051
Флухинконазол / Fluhinconazole	3.526	3.526	2.713	3.526	3.522	3.621	3.557	0.000	-0.813	0	-0.004	0.095	0.031
Фуберидазол / Fuberidazole	2.677	2.339	2.531	2.683	2.085	2.648	2.472	-0.338	-0.146	0.006	-0.592	-0.029	-0.205
Фурконазол-цис Furconazole-cis	2.945	2.436	2.600	2.947	2.516	2.771	2.745	-0.509	-0.345	0.002	-0.429	-0.174	-0.2
Ципроконазол / Ciproconazole	2.689	2.441	2.497	2.683	2.281	2.666	2.544	-0.248	-0.192	-0.006	-0.408	-0.023	-0.145
Эпоксиконазол / Eproxiconazole	2.019	2.081	2.222	2.019	2.073	2.385	2.159	0.062	0.203	0	0.054	0.366	0.14
Этаконазол / Etaconazole	2.388	2.591	2.650	2.388	2.568	2.758	2.572	0.203	0.262	0	0.18	0.37	0.184
Этридиазол / Ethridiazole	2.39	2.300	2.333	2.390	2.377	2.680	2.482	-0.090	-0.057	0	-0.013	0.29	0.092
<i>Карбаматы / Carbamates</i>													
Алдикарб / Aldicarb	5.464	5.688	4.656	5.464	5.467	5.459	5.577	0.224	-0.808	0	0.003	-0.005	0.113
Аминокарб / Aminocarb	3.841	3.74	3.995	3.847	3.841	3.834	3.79	-0.101	0.154	0.006	-0	-0.007	-0.051
Бутилат / Butylate	1.764	2.046	2.384	1.764	1.768	1.829	1.905	0.282	0.62	0	0.004	0.065	0.141
Вернолат / Vernolat	2.181	2.182	2.184	2.181	1.986	2.179	2.085	0.001	0.003	0	-0.195	-0.002	-0.096
Диаллат / Diallat	2.835	2.993	2.88	2.81	2.838	2.811	2.915	0.158	0.045	-0.025	0.003	-0.024	0.08
Изолан / Isolan	4.291	4.256	3.995	4.291	4.283	4.289	4.269	-0.035	-0.296	0	-0.008	-0.002	-0.022
Карбанилат / Carbanilate	3.853	3.949	3.936	3.847	3.857	3.962	3.905	0.096	0.083	-0.006	0.004	0.109	0.052
Карбарил / Carbaryl	2.729	3.171	3.014	2.729	2.87	2.722	3.021	0.442	0.285	0	0.141	-0.007	0.292
Карбосульфан / Carbosulfan	3.725	3.551	4.076	3.724	3.407	3.732	3.479	-0.173	0.352	0	-0.317	0.008	-0.245
Карбофуран / Carbofuran	4.646	4.613	4.055	4.646	4.641	4.722	4.627	-0.033	-0.591	0	-0.005	0.076	-0.019

Продолжение таблицы 2 на стр. 821.

Окончание таблицы 2. Начало на стр. 818.

Вещество Compound name	lg (1/DL ₅₀), экспериментальное lg (1/DL ₅₀), experimental	lg (1/DL ₅₀), прогнозируемое lg (1/DL ₅₀), predicted						Предел ошибки Margin of error					
		ЛР LR	к-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM	ЛР LR	к-БС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM
<i>Карбаматы / Carbamates</i>													
Молилат / Molilat	2.647	2.114	2.339	2.647	2.092	2.759	2.102	-0.533	-0.308	0	-0.555	0.112	-0.545
Пебулат / Pebulat	2.302	2.033	2.184	2.301	2.089	2.293	2.062	-0.268	-0.117	0	-0.212	-0.008	-0.239
Пиримикарб / Pyrimicarb	3.206	2.785	2.37	3.206	2.345	3.194	2.565	-0.421	-0.836	0	-0.861	-0.012	-0.641
Пропамоккарб гидрохлорид Propramocarb hydrochloride	2.139	2.415	2.441	2.139	2.611	2.12	2.515	0.276	0.302	0	0.472	-0.019	0.376
Пропоксур / Propoxur	3.665	3.208	3.416	3.665	3.148	3.656	3.179	-0.457	-0.249	0	-0.517	-0.009	-0.486
Тиомочевина / Thiourea	2.785	2.693	2.868	2.81	2.784	2.877	2.735	-0.092	0.083	0.025	-0.001	0.092	-0.05
Тирам / Tiram	2.126	2.265	2.203	2.126	2.133	2.123	2.198	0.139	0.077	0	0.007	-0.003	0.072
Феноксикарб / Phenoxycarb	1.366	1.614	2.028	1.366	1.659	1.333	1.636	0.248	0.662	0	0.293	-0.033	0.27
Фурациокарб / Furatiocarb	3.858	3.694	4.349	3.858	3.866	3.9	3.779	-0.164	0.491	0	0.008	0.042	-0.079
Хлорпрофам / Chlorprofam	1.979	2.376	2.594	1.979	2.606	2.056	2.492	0.397	0.615	0	0.627	0.077	0.513
Циклоат / Cycloat	2.07	2.087	2.339	2.07	2.073	2.104	2.079	0.017	0.269	-0	0.003	0.034	0.009
<i>Фосфорорганические соединения (ФОС) / Organophosphorus compounds (OPs)</i>													
Анилофос / Anilophos	2.892	2.934	3.537	2.906	2.859	2.902	2.881	0.042	0.646	0.015	-0.032	0.01	-0.011
Дикротофос / Dicrotophos	4.203	4.281	4.161	4.203	4.209	4.236	4.222	0.078	-0.042	0	0.006	0.033	0.019
Диметоат / Dimethoate	2.971	3.223	3.201	2.978	3.515	3.026	3.27	0.252	0.23	0.007	0.544	0.054	0.299
Димефокс / Dimefox	5.188	4.784	4.458	5.176	5.182	5.248	5.215	-0.404	-0.73	-0.012	-0.005	0.06	0.027
Дихлоровас / Dichlorovas	3.778	4.069	3.722	3.778	3.773	3.899	3.836	0.292	-0.056	0	-0.005	0.121	0.058
Кадусафос / Kadousafos	3.953	3.812	4.253	3.953	3.95	3.93	3.94	-0.141	0.3	-0	-0.003	-0.024	-0.013
Малатион / Malathion	2.394	2.289	2.885	2.394	2.391	2.395	2.393	-0.105	0.491	-0	-0.003	0.001	-0.001
Метакрифос / Metacrifos	2.549	2.907	2.853	2.549	2.687	2.599	2.643	0.357	0.304	0	0.138	0.05	0.094
Мипафокс / Mirafox	3.482	3.574	3.88	3.482	3.487	3.479	3.483	0.092	0.397	0	0.005	-0.004	0
Оксидеметон-метил Oxudemetone-methyl	3.71	3.568	3.201	3.71	3.709	3.679	3.694	-0.142	-0.509	0	-0.001	-0.031	-0.016
Паратион-метил Methyl-Parathion	4.943	4.245	4.364	4.943	4.182	4.949	4.565	-0.699	-0.58	0	-0.761	0.006	-0.378
Паратион-этил Ethyl-parathion	5.163	4.859	5.243	5.176	4.86	5.34	5.1	-0.304	0.08	0.012	-0.303	0.177	-0.063
Протиофос / Protiophos	2.584	2.689	2.885	2.584	2.847	2.733	2.79	0.105	0.301	0	0.263	0.149	0.206
Тербуфос / Terbufos	5.301	5.641	5.243	5.283	5.623	5.354	5.489	0.34	-0.058	-0.018	0.322	0.053	0.188
Трихлорфон / Trichlorophone	2.921	2.517	3.201	2.906	2.921	3.018	2.969	-0.404	0.28	-0.015	0	0.097	0.049
Фенитротрион / Fenitrotion	2.985	3.865	3.682	2.978	3.785	3.044	3.415	0.88	0.697	-0.007	0.801	0.06	0.43
Фентион / Fention	3.118	3.257	3.537	3.118	3.118	3.081	3.1	0.139	0.419	0	0	-0.037	-0.018
Форат / Forat	5.265	5.173	5.243	5.283	5.261	5.397	5.329	-0.092	-0.022	0.018	-0.004	0.132	0.064
Фосфамидон / Phosphamidon	4.603	3.715	3.975	4.603	3.514	4.708	4.111	-0.888	-0.628	0	-1.088	0.105	-0.491
Хлорпирифос / Chlorpyrifos	3.678	4.279	3.458	3.678	4.229	3.687	3.958	0.601	-0.22	0	0.551	0.009	0.28

Примечание. ЛР – линейная регрессия; к-БС – к-ближайших соседей; ДПР – деревья принятия решений; МОП – метод опорных векторов; НС – нейронная сеть; АМ – ансамблевые методы.

Note: LR – linear regression; k-NN – k-nearest neighbours; DT – decision trees; SVM – support vector machine; NN – neural network; EM – ensemble methods.

Таблица 3 / Table 3

Результаты внутренней 10-кратной кросс-валидации математических моделей
Results of internal 10-fold cross-validation of mathematical models

		ЛР LR	к-НС k-NN	ДПР DT	МОП SVM	НС NN	АМ EM	
<i>Хлороорганические соединения (ХОС) / Organochlorine compounds (OCs)</i>								
Тренировочный набор Training set	RMSE	0.2726	0.3593	0.1975	0.2112	0.2693	0.2214	
	r^2	0.888	0.821	0.941	0.933	0.936	0.937	*НС + МОВ / NN + SVM
10-кратная кросс-валидация 10-fold validation	RMSE	0.7491	0.5766	0.8271	0.5055	0.5636	0.4988	
	r^2	0.166	0.511	0.145	0.632	0.628	0.659	*НС + МОВ / NN + SVM
<i>Азолы / Azoles</i>								
Тренировочный набор Training set	RMSE	0.2324	0.2612	0.0026	0.2434	0.2495	0.1351	
	r^2	0.742	0.683	1	0.725	0.819	0.923	НС + МОВ + ДПР / NN + SVM + DT
10-кратная кросс-валидация 10-fold validation	RMSE	0.4418	0.3713	0.5112	0.3866	0.4937	0.3542	
	r^2	0.306	0.361	0.18	0.419	0.408	0.476	*НС + МОВ + ДПР / NN + SVM + DT
<i>Карбаматы / Carbamates</i>								
Тренировочный набор Training set	RMSE	0.2668	0.4263	0.0079	0.3253	0.0508	0.2839	
	r^2	0.934	0.845	1	0.905	0.998	0.926	*ЛР + МОВ / LR + SVM
10-кратная кросс-валидация 10-fold validation	RMSE	0.7121	0.682	0.9612	0.6307	0.8718	0.6229	
	r^2	0.58	0.571	0.343	0.641	0.456	0.651	*ЛР + МОВ / LR + SVM
<i>Фосфорорганические соединения (ФОС) / Organophosphorus compounds (OPs)</i>								
Тренировочный набор Training set	RMSE	0.4079	0.4164	0.0085	0.4056	0.0792	0.2049	
	r^2	0.824	0.842	1	0.826	0.997	0.958	*НС + МОВ / NN + SVM
10-кратная кросс-валидация 10-fold validation	RMSE	0.8256	0.7571	0.9711	0.6979	0.7922	0.6936	
	r^2	0.336	0.423	0.268	0.504	0.472	0.526	*НС + МОВ / NN + SVM

При изучении табл. 2 можно предположить, что алгоритм дерева принятия решений является наиболее верным при прогнозировании пероральной токсичности веществ всех изучаемых групп из-за наименьших значений пределов ошибок. Однако данное предположение не совсем верно, поскольку нельзя делать выводы о лучшей модели для прогнозирования без изучения пригодности и надёжности модели. Так, была проведена внутренняя валидация полученных моделей (табл. 3).

В табл. 3 представлены параметры качества построенных моделей – RMSE, r^2 . Модель QSAR приемлема для использования, когда она имеет значение $r^2 > 0,6$ для тренировочного набора и $r^2 > 0,5$ – при проведении кросс-валидации (в табл. 3 жирным шрифтом отмечены те значения, которые соответствуют данным условиям).

Следует отметить, что значения коэффициентов детерминации во всех тренировочных наборах составляют от 0,683 до 1, то есть $r^2 > 0,6$, что гарантирует высокую степень соответствия моделей QSAR использованным данным. Диапазон значений коэффициента детерминации при проведении 10-кратной кросс-валидации значительно ниже (см. табл. 3).

Обсуждение

С учётом параметров, необходимых для выбора модели, установлено следующее:

1) для группы хлороорганических соединений метод опорных векторов показал наилучший результат внутренней прогностической способности разработанной модели: 63,2% случаев изменения дескрипторов приводили к изменению значения конечной точки;

2) для группы азолов не удалось создать модель, которая соответствовала бы всем необходимым требованиям;

3) для группы карбаматов метод опорных векторов показал наилучший результат внутренней прогностической

способности разработанной модели: 64,1% случаев изменения дескрипторов приводили к изменению значения конечной точки;

4) для группы фосфорорганических соединений метод опорных векторов показал наилучший результат внутренней прогностической способности разработанной модели: 50,4% случаев изменения дескрипторов приводили к изменению значения конечной точки.

Низкие значения RMSE моделей прогнозирования токсичности ХОС, карбаматов и ФОС по сравнению с другими указывают на то, что разработанные модели QSAR стабильны при прогнозировании неизвестных соединений в тестовом наборе.

Несмотря на то что для трёх групп веществ основной метод прогнозирования – метод опорных векторов, нужно отметить, что для каждой выборки веществ были сгенерированы собственные дескрипторы. Поэтому объединение соединений из разных групп в одну выборку для прогнозирования не допускается.

Для решения вопроса прогнозирования пероральной токсичности азолов дополнительно были использованы ансамблевые модели, решающие проблему небольшого размера выборки за счёт усреднения и включения нескольких моделей. При оценке модели, состоящей из нейронных сетей, опорных векторов и деревьев принятия решения, было получено, что $r^2 = 0,923$ (для тренировочного набора) и $r^2 = 0,476$ (при кросс-валидации). Данные значения лучше тех значений, которые были получены в «классических» моделях регрессии, но тем не менее статистические данные всё ещё не соответствуют требованиям, предъявляемым выше, поэтому модель не может быть рекомендована для использования.

При применении ансамблевых методов с различными комбинациями моделей на других группах пестицидов был замечен рост коэффициента детерминации при перекрёстной проверке и уменьшение значений RMSE. Для прогнозирования ХОС и ФОС оптимально использование модели,

в которой происходит комбинирование нейронных сетей и метода опорных векторов, для карбаматов – ансамблевой модели, включающей в себя линейную регрессию и метод опорных векторов (см. табл. 3).

Ограничения исследования. Исследование ограничено количеством исследуемых соединений, классами химических соединений, областью распространения полученных в ходе моделирования результатов.

Заключение

Целью данной работы было создать и сравнить математические модели для прогнозирования острой пероральной токсичности различных классов пестицидов. В исследовании было включено 4 класса пестицидов (ХОС, азолы, карбаматы, ФОС) с общим количеством 100 соединений с дескрипторами, рассчитанными программным обеспечением PaDEL-Descriptors ver. 2.21. В программе WEKA были

построены модели регрессии, подвергнутые процедуре внутренней валидации. Была произведена оценка качества полученных моделей с помощью статистических параметров RMSE, r^2 . С учётом возможностей ансамблевых методов моделирования было установлено, что для прогнозирования острой пероральной токсичности ХОС и ФОС оптимально использование модели, в которой происходит комбинирование нейронных сетей и метода опорных векторов, для карбаматов – ансамблевой и включающей в себя линейную регрессию и метод опорных векторов. Для веществ из группы азолов не удалось создать модели, которая соответствовала бы необходимым требованиям ($r^2 > 0,6$ для тренировочного набора и $r^2 > 0,5$ при проведении кросс-валидации). Для решения данной проблемы необходимо вычленив из данного класса класс триазолы для улучшения производительности моделей, а также повторно оценить адекватность результатов экспериментальных исследований, полученных из источников литературы.

Литература

(п.п. 1–2, 4–21 см. References)

3. Руководство Р 1.2.3156-13. Оценка токсичности и опасности химических веществ и их смесей для здоровья человека. М.; 2014.

References

- Hamadache M., Benkortbi O., Hanini S., Amrane A., Khaouane L., Si Moussa C. A quantitative structure activity relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction. *J. Hazard. Mater.* 2016; 303: 28–40. <https://doi.org/10.1016/j.jhazmat.2015.09.021>
- Wang Y., Ning Z.H., Tai H.W., Long S., Qin W.C., Su L.M., et al. Relationship between lethal toxicity in oral administration and injection to mice: Effect of exposure routes. *Regul. Toxicol. Pharmacol.* 2015; 71(2): 205–12. <https://doi.org/10.1016/j.yrtph.2014.12.019>
- Manual R 1.2.3156-13. Assessment of toxicity and danger of chemicals and their mixtures for human health. Moscow; 2014. (in Russian)
- Schrage A., Hempel K., Schulz M., Kollé S.N., van Ravenzwaay B., Landsiedel R. Refinement and reduction of acute oral toxicity testing: a critical review of the use of cytotoxicity data. *Altern. Lab. Anim.* 2011; 39(3): 273–95. <https://doi.org/10.1177/026119291103900311>
- Burden N., Mahony C., Müller B.P., Terry C., Westmoreland C., Kimber I. Aligning the 3Rs with new paradigms in the safety assessment of chemicals. *Toxicol.* 2015; 330: 62–6. <https://doi.org/10.1016/j.tox.2015.01.014>
- CAS REGISTRY®. Available at: <https://www.cas.org/cas-data/cas-registry>
- Carrio P., Sanz F., Pastor M. Towards a unifying strategy for predicting toxicological endpoints based on structure. *Arch. Toxicol.* 2016; 90: 2445–60. <https://doi.org/10.1007/s00204-015-1618-2>
- Reyes A.B., Bayich V.B. In silico toxicology: computational methods for predicting chemical toxicity. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2016; 6(2): 147–72. <https://doi.org/10.1002/wcms.1240>
- Villaverde J.J., Sevilla-Moran B., Lopez-Goti S., Alonso-Prados J.L., Sandin-España P. QSAR/QSPR models based on quantum chemistry for assessing the risk of pesticides in accordance with current European legislation. *SAR QSAR Environ. Res.* 2019; 31(1): 49–72. <https://doi.org/10.1080/1062936x.2019.1692368>
- Organization for Economic Cooperation and Development (OECD). Guidance Document on the Validation of (Quantitative) Structure–activity Relationships [(Q)SAR] Models. ENV/JM/MONO2; 2007. Available at: <https://www.oecd.org/env/guidance-document-on-the-validation-of-quantitative-structure-activity-relationship-q-sar-models-9789264085442-en.htm>
- Gramatica P. On the development and validation of QSAR models. *Methods Mol. Biol.* 2013; 930: 499–526. https://doi.org/10.1007/978-1-62703-059-5_21
- Spiegel J., Senderowitz H. Evaluation of QSAR equations for virtual screening. *Int. J. Mol. Sci.* 2020; 21(21): 7828. <https://doi.org/10.3390/ijms21217828>
- Achary P.G.R. Applications of quantitative structure-activity relationships (QSAR) based virtual screening in drug design: a review. *Mini Rev. Med. Chem.* 2020; 20(14): 1375–88. <https://doi.org/10.2174/1389557520666200429102334>
- Kim K.H., Kabir E., Jahan S.A. Exposure to pesticides and the associated human health effects. *Sci. Total Environ.* 2017; 575: 525–35. <https://doi.org/10.1016/j.scitotenv.2016.09.009>
- Lewis K.A., Tzilivakis J., Warner D., Green A. An international database for pesticide risk assessments and management. *Hum. Ecol. Risk Assess.* 2016; 22(4): 1050–64.
- CDC. Registry of Toxic Effects of Chemical Substances (RTECS); 2022. Available at: <http://www.cdc.gov/niosh/rtecs>
- Dearden J.C., Cronin M.T.D., Kaiser K.L.E. How not to develop a quantitative structure–activity or structure–property relationship (QSAR/QSPR). *SAR QSAR Environ. Res.* 2009; 20(3–4): 241–66. <https://doi.org/10.1080/10629360902949567>
- Yap C.W. PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *J. Computat. Chem.* 2011; 32(7): 1466–74. <https://doi.org/10.1002/jcc.21707>
- Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I.H. The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter.* 2009; 11(1): 10–8.
- Golbraikh A., Tropsha A. Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. *J. Comput. Aided Mol. Des.* 2002; 16(5/6): 357–69. <https://doi.org/10.1023/a:1020869118689>
- Frimayanti N., Yam M.L., Lee H.B., Othman R., Zain S.M., Rahman N.A. Validation of quantitative structure-activity relationship (QSAR) Model for photosensitizer activity prediction. *Int. J. Mol. Sci.* 2011; 12(12): 8626–44. <https://doi.org/10.3390/ijms12128626>