

**ЭКСПЕРТНАЯ МОДЕЛЬ ОЦЕНКИ ОСТРОЙ ТОКСИЧНОСТИ СТРУКТУРНО
РОДСТВЕННЫХ СОЕДИНЕНИЙ - ПРОИЗВОДНЫХ АЗАТРИЦИКЛОНОНА**

Лапко Е.Ю., Злобин А.В., Лапшинов О.В., Левченко Е.В., Румянцева В.И.

Федеральное государственное бюджетное учреждение «27 Научный центр»

Министерства обороны Российской Федерации, г. Москва, Бригадирский переулок, д. 13,

8(499) 261-11-44, 27nc_2@mil.ru

Резюме. В данной работе рассмотрена возможность прогнозирования токсических свойств соединений азабрэнданового типа по их молекулярной структуре. На основе синтезированных соединений была создана нейро-нечеткая модель «структура – активность». Для прогнозирования свойств данных соединений, был использован такой численный метод исследования, как нечеткая логика и показана возможность применения экспертной модели на основе алгоритма Сугено для определения острой токсичности производных 4-азатрицикло[4.2.1.0^{3,7}]нонанов по структурной формуле. Предложенная методология создания экспертной модели может применяться для любых классов соединений при условии подбора соответствующих дескрипторов, характеризующих особенности молекулярной структуры выбранного класса веществ, и наличии достаточного количества соединений для обучения нейро-нечеткой гибридной сети.

Ключевые слова: нейро-нечеткое моделирование, прогноз, обучающая выборка, дескриптор, острая токсичность.

**EXPERT MODEL FOR EVALUATING THE ACUTE TOXICITY OF
STRUCTURALLY RELATED COMPOUNDS - AZATRICYCLONONANE DERIVATIVES**

Lapko E.Yu., Zlobin A.V., Lapshinov O.V., Levchenko E.V., Rumyantseva V.I.

"27 Central research Institute" of the Ministry of defense of the Russian Federation

Abstract. In this paper, the possibility of predicting the toxic properties of azabrendan-type compounds by their molecular structure is considered. Based on the synthesized compounds, a neuro-fuzzy model “structure - activity” was created. To predict the properties of these compounds, a numerical research method such as fuzzy logic was used and the possibility of using an expert model based on the Sugeno algorithm to determine the acute toxicity of 4-azatricyclo [4.2.1.0^{3,7}] nonanes derivatives by the structural formula was shown. The proposed methodology for creating an expert model can be applied to any classes of compounds, provided that appropriate descriptors are selected that characterize the molecular structure of the selected class of substances and that there are enough compounds to train a neuro-fuzzy hybrid network.

Keywords: neuro fuzzy modeling, prediction, training sample, descriptor, acute toxicity.

В настоящее время известно более 19 миллионов химических соединений, из которых 60-80 тысяч производится в промышленном масштабе. При этом предусмотренную законодательством экологическую экспертизу прошли не более полутора тысяч, а понятием класс опасности охвачено чуть более тысячи веществ [1]. Каждый год появляются новые органические соединения, экологическая опасность которых неизвестна. Возможность прогнозирования токсических свойств соединений по их молекулярной структуре позволит осуществить комплекс превентивных мер по организации их безопасного производства и использования, а также планирования мероприятий ликвидации последствий чрезвычайных ситуаций на объектах по их производству.

Внимание исследователей наряду с методологией нейросетевого моделирования для прогнозирования свойств органических соединений, привлекает такой численный метод исследования, как нечеткая логика. В данной работе предложена модель экспертной оценки острой токсичности замещенных 4-азатрицикло[4.2.1.0^{3,7}]нонанов на основе алгоритма Сугено, созданная с использованием нейро-нечетких гибридных сетей.

Трициклические структуры окса- и азабрэнданов, а также их карбонилсодержащие аналоги представляют интерес как потенциальные биологически активные системы или синтоны для их получения. В то время как свойства оксабрэнданов достаточно подробно исследованы, замещенные 4-азатрицикло[4.2.1.0^{3,7}]нонаны (азабрэнданы), ввиду малой синтетической доступности, изучены недостаточно.

Кроме того, интерес к исследованию свойств 4-азатрицикло[4.2.1.0^{3,7}]нонанов с использованием нейро-нечетких гибридных сетей обусловлен:

- наличием в молекулах фрагментов, родственных бициклическим соединениям, что объясняет наличие широкого спектра биологической активности;
- жесткостью молекул с «закрепленными» в пространстве заместителями, превращающих их в удобные модели для изучения количественной зависимости «структура – активность».

В ходе синтетических исследований изучались различные способы получения соединений азабрэнданового типа с заместителями в различном положении каркаса и модификации структуры полученных веществ. Оптимальные результаты, характеризующиеся высокими выходами и чистотой продукта, были достигнуты при использовании циклизации непредельных бициклических структур, содержащих аминогруппу. Синтезированные

целевые соединения были использованы для формирования учебной и внутренней контрольных выборок с целью создания нейро-нечеткой модели «структура – активность».

Процесс нейро-нечеткого моделирования зависимости «структура-свойство» осуществляли путем последовательного выполнения следующих операций:

- определение размеров обучающей и контрольной выборки;
- описание структуры исследуемых соединений числовыми параметрами - дескрипторами;
- выбор информативных дескрипторов;
- выбор параметров обучения гибридной нейросети;
- обучение адаптивной нейро-нечеткой сети и оценка результатов этого обучения;
- выбор параметров для оценки качества обучения нейросети;
- использование полученной гибридной модели для прогноза свойств и активностей неизвестных соединений.

Моделирование структур молекул выполняли на базе программ CambridgeSoft ChemBio Office Ultra 13.0 [2] с 3D оптимизацией геометрии молекул методом молекулярной механики. Для вычисления молекулярных дескрипторов использовали программный комплекс DRAGON 7 [3]. На основе смоделированных молекулярных структур были рассчитаны 22 класса дескрипторов [4] и получено универсальное множество значений молекулярных описателей.

Далее был осуществлен предварительный подбор параметров (дескрипторов), характеризующих структуру молекул исследуемых веществ. Исходя из анализа литературных источников, информативности и значений коэффициента корреляции рассмотренных значений дескрипторов с активностью, в качестве наиболее значимых были отобраны коэффициент распределения «октанол-вода» LogP, энергия HOMO, 3D молекулярные дескрипторы: J3D (3D-Balaban index), SPAN (span R, DISPM d COMMA2 value / weighted by atomic masses), Mor09u (3D-MoRSE - signal 09 / unweighted), Mor19u (3D-MoRSE - signal 19 / unweighted), Mor31m (3D-MoRSE - signal 31 / weighted by atomic masses), Mor17e (3D-MoRSE - signal 17 / weighted by atomic Sanderson electronegativities), Mor19e (3D-MoRSE - signal 19 / weighted by atomic Sanderson electronegativities), H0e (H autocorrelation of lag 0 / weighted by atomic Sanderson electronegativities), Ds (D total accessibility index / weighted by atomic electrotopological states) [5]. Полученные величины были использованы в качестве входящих параметров разрабатываемой искусственной нейро-нечеткой гибридной модели. Задача прогнозирования реализована в среде разработки MATLAB Fuzzy Logic Toolbox.

На основе подхода, предложенного Баскиным И.И. [6], проведено нормирование входных параметров нейросети таким образом, чтобы их значения были от 0 до 1. Наибольшее значение дескриптора из исследуемой выборки и принимали равным 1, а остальные значения получали как частное от деления на него.

Для моделирования зависимости «структура - свойство» рассматривали нейро-нечеткие гибридные сети, которые входят в состав пакета Fuzzy logic программы MATLAB R 2011a. В процессе обучения осуществляли контроль прогнозирующей способности нейро-нечеткой гибридной сети при помощи заранее выбранной контрольной выборки примеров, что позволило избежать «переобучения».

Для построения модели использовали от 6 до 8 различных выбранных дескрипторов, подаваемых на вход нейро-нечеткой гибридной сети. Обучение сетей осуществляли выбирая функции принадлежности, а также изменяя количество эпох (однократное представление всех обучающих входных данных на входные нейроны сети). Функция принадлежности является одним из основных параметров для входной и выходной переменных. Для первой переменной была использована функция типа *gaussmf*, для второй - *const*. Выбор функции *gaussmf* обусловлен предположением о нормальном распределении случайной величины, которое определяется гауссовской функцией по теории вероятности. Функция *const* выбрана из-за нулевого порядка алгоритма Сугено [7].

Для моделирования использовали трехвыборочный метод и делили набор данных на 3 выборки: обучающую, внутреннюю контрольную и внешнюю контрольную. Обучающая выборка использовалась для построения моделей. При помощи внутренней контрольной выборки осуществляли контроль прогнозирующей способности сети, что позволило избежать «переобучения» [8]. Также внутреннюю контрольную выборку использовали в целях отбора оптимальной для прогнозирования модели. Ошибку прогнозирования на внешней контрольной выборке, которая не участвовала в построении и отборе модели, использовали для оценки прогнозирующей способности модели.

Учебная выборка должна соответствовать определенным требованиям, одно из которых заключается в следующем: минимальное количество объектов в обучающей выборке должно быть в 3-5 раз больше, чем число описателей. Поэтому была создана база соединений, сформированная исходя из структурных особенностей изучаемых веществ. Для учебной выборки использовали вещества, фрагментарно присутствующие в каркасе молекулы азабрэндана. При создании массива данных выборок использовали значения среднесмертельных токсодоз, взятых из литературных источников и баз данных, а для вновь синтезированных соединений значения определяли экспериментально.

Прогнозирующую способность моделей оценивали с помощью коэффициента перекрестной оценки q^2 , предложенного Р.Крамером [9]. Характеристики нейро-нечетких гибридных моделей, показавших наилучшую прогнозирующую способность, представлены в таблице 1.

Таблица 1 - Характеристики нейро-нечетких гибридных моделей, полученных при моделировании зависимости «структура – активность» для 4-азатрициклических соединений

Шифр модели	Ds	HOe	HOMO	LOGP	J3D	Mor19u	Mor19e	SPAN	DISPm	Mor31m	Mor17e	Mor09u	Количество входных переменных	Число созданных правил	q^2
z11 6	+	+	+	+	+	+							6	729	0,760
z12 6	+	+	+	+	+					+			6	729	0,750
z13 6	+	+	+	+	+						+		6	729	0,651
z14 6	+	+	+	+	+		+						6	729	0,729
z15 6	+	+	+	+	+							+	6	729	0,707
z16 6	+	+	+	+	+			+					6	729	0,712
z17 6	+	+	+	+	+				+				6	729	0,748
z21 7	+	+	+	+	+	+				+			7	2187	0,781
z22 7	+	+	+	+	+	+					+		7	2187	0,792
z23 7	+	+	+	+	+	+	+						7	2187	0,823
z24 7	+	+	+	+	+	+						+	7	2187	0,728
z25 7	+	+	+	+	+	+		+					7	2187	0,737
z26 7	+	+	+	+	+	+			+				7	2187	0,775
z31 8	+	+	+	+	+	+	+			+			8	6561	0,873
z32 8	+	+	+	+	+	+	+				+		8	6561	0,847
z33 8	+	+	+	+	+	+	+					+	8	6561	0,896
z34 8	+	+	+	+	+	+	+	+					8	6561	0,922
z35 8	+	+	+	+	+	+	+		+				8	6561	0,913

Из полученных моделей была выбрана гибридная сеть (z34|8) на основе 8 молекулярных дескрипторов (Ds, HOe, HOMO, LogP, J3D, Mor19u, Mor19e, SPAN), для входной переменной использующая функцию принадлежности типа *gaussmf*, для выходной переменной - *const*, показавшая наиболее достоверные результаты прогноза для соединений внутренней контрольной выборки. Данная нейро-нечеткая гибридная модель была использована для прогнозирования значений острой токсичности 4-азатрицикло[4.2.1.0^{3,7}]нонанов. Структура модели представлена на рисунке 1.

Прогнозирующие способности данной гибридной сети проверены с помощью внешней контрольной выборки соединений 4-азатрициклической структуры. Расчетные и экспериментальные значения среднесмертельных доз соединений представлены в таблице 2.

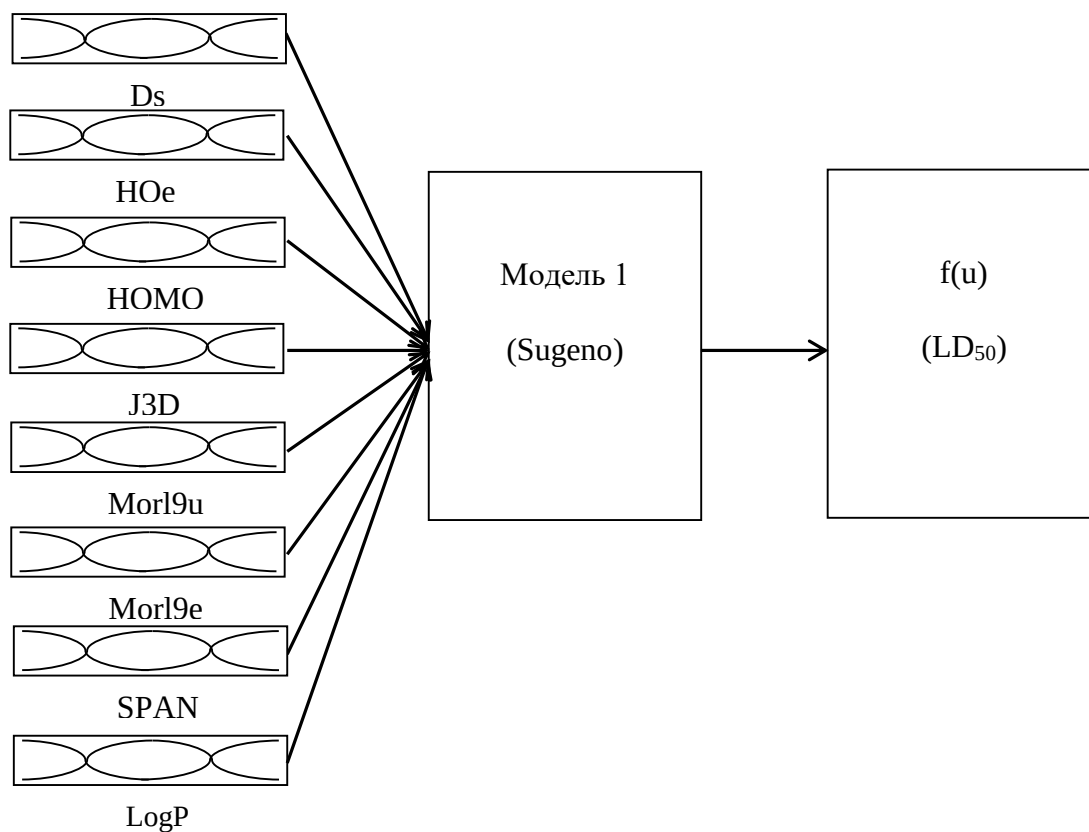
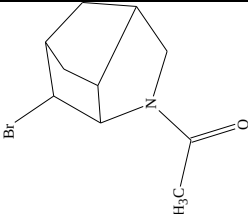
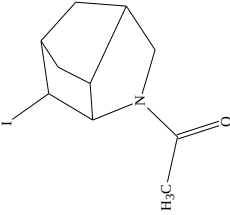
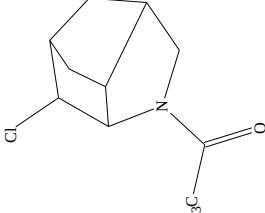


Рисунок 1 - Структура нейро-нечеткой гибридной модели, использованная для прогнозирования значений острой токсичности 4-азатрицикло[4.2.1.0^{3,7}]нонанов

Таблица 2 - Экспериментальные и расчетные значения LD₅₀ для соединений 4-азатрициклической структуры (нейро-нечеткая гибридная модель z34|8)

Структурная формула соединения	LD ₅₀ , мг/кг	
	экспериментальная*	расчетная
	105 ± 42 n = 30	138
	180 ± 35 n = 30	206
	70 ± 28 n = 30	95
Примечание – *Достоверность p = 0,05; n – количество лабораторных животных.		

Прогнозные значения укладываются в доверительные интервалы экспериментально полученных среднесмертельных доз, что говорит об удовлетворительных прогностических возможностях модели z34|8 по отношению к 4-азатрициклическим соединениям. Полученные значения острой токсичности при помощи компилятора MATLAB были экспортированы в язык программирования C++, и на основе алгоритма классификации была написана программа определения класса опасности вещества.

Таким образом, созданная экспертная система позволяет определять острую токсичность соединений ряда 4-азатрицикло[4.2.1.0^{3,7}]нонанов по структурной формуле. А предложенная методология создания экспертной модели может применяться для любых классов соединений при условии подбора соответствующих дескрипторов, характеризующих особенности молекулярной структуры выбранного класса веществ, и наличии достаточного количества соединений для обучения нейро-нечеткой гибридной сети.

Список литературы

1. *Дербишер Е.В., Веденина Н.В., Александрина А.Ю., Дербишер В.Е.* Методология прогнозирования класса опасности малоизученных органических соединений / Современные наукоемкие технологии. 2007. № 8. С. 60-62.
2. CambridgeSoft ChemBio Office Ultra 13.0 [Electronic resources]. [http:// www.cambridgeSoft-.com](http://www.cambridgeSoft-.com) / Date of acces: 21.08.2016.
3. Dragon 7 [Electronic resources]. <http://www.taletе.mi.it> / Date of acces: 12.02.2020.
4. *Todeschini R., Consonni V.* Handbook of Molecular Descriptors / Weinheim: WILEY-WCH Publishers, 2000. 688 p.
5. *Balaban A.T.* From Chemical Topology to 3D Geometry / J.Chem.Inf.ComputSci, 1997. № 37. P. 645-650.
6. *Баскин И.И., Гальберштам Н.М., Палюлин В.А.* Компьютерная реализация искусственных нейронных сетей для решения задач по выявлению связи «структура-свойство» / Информационные технологии, 1997. № 9. С. 27-30.
7. *Круглов В.В.* Нечеткая логика и искусственные нейронные сети / М.: Издательство Физико-математической литературы, 2001. 224 с.
8. *Baskin I.I., Skvortsova M.I., Palyulin V.A., Zefirov N.S.* Quantitative chemical structure - property/activity relationship studies using artificial neural networks. / Foundations of Computing and Decision Sciences, 1997. № 22. № 2. P. 107-116.
9. *Cramer R.D., Patterson D.E., Bunce J.D.* Comparative molecular field analysis (CoMFA). Effect of shape on binding of steroids to carrier proteins / Hi. Am.Chem. Soc, 1988. № 110. P. 5959-5967.

References

1. *Derbisher E.V., Vedenina N.V., Aleksandrina A.Yu., Derbisher V.E.* Methodology for predicting the hazard class of little-known organic compounds / Modern high technology. 2007. № 8. P. 60-62. [In Russian].
2. CambridgeSoft ChemBio Office Ultra 13.0 [Electronic resources]. [http:// www.cambridgeSoft-.com](http://www.cambridgeSoft-.com) / Date of acces: 21.08.2016.
3. Dragon 7 [Electronic resources]. <http://www.taletе.mi.it> / Date of acces: 12.02.2020.
4. *Todeschini R., Consonni V.* Handbook of Molecular Descriptors / Weinheim: WILEY-WCH Publishers, 2000. 688 p.
5. *Balaban A.T.* From Chemical Topology to 3D Geometry / J.Chem.Inf.ComputSci, 1997. № 37. P. 645-650.

6. *Baskin I.I., Gal'bershtam N.M., Palyulin V.A.* Computer implementation of artificial neural networks to solve the problems of revealing the connection «structure-property» / Information technology, 1997. № 9. P. 27-30. [In Russian].

7. *Kruglov V.V.* Fuzzy logic and artificial neural networks / М.: Publishing house of physical-mathematical literature, 2001. 224 p. [In Russian].

8. *Baskin I.I., Skvortsova M.I., Palyulin V.A., Zefirov N.S.* Quantitative chemical structure - property/activity relationship studies using artificial neural networks. / Foundations of Computing and Decision Sciences, 1997. № 22. № 2. P. 107-116.

9. *Cramer R.D., Patterson D.E., Bunce J.D.* Comparative molecular field analysis (CoMFA). Effect of shape on binding of steroids to carrier proteins / *Hi. Am.Chem. Soc*, 1988. № 110. P. 5959-5967.