



DOI <https://doi.org/10.11603/2312-0967.2023.3.13985>

УДК 615.011.4:004.032.26

## ПРОГНОЗУВАННЯ СТУПЕНЯ НЕБЕЗПЕЧНОСТІ/РИЗИКУ ЗАЛИШКОВИХ РОЗЧИННИКІВ У ЛІКАРСЬКИХ ЗАСОБАХ МЕТОДАМИ ХЕМОМЕТРІЇ

Я. М. Пушкарьова, А. В. Калюженко

Національний медичний університет імені О. О. Богомольця, Київ

[yaroslava.pushkarova@gmail.com](mailto:yaroslava.pushkarova@gmail.com)

### ІНФОРМАЦІЯ

Надійшла до редакції / Received:  
24.07.2023

Після доопрацювання / Revised:  
07.09.2023

Прийнято до друку / Accepted:  
14.09.2023

### Ключові слова:

залишковий розчинник;  
класифікація;  
дескриптор;  
штучна нейронна мережа.

### АНОТАЦІЯ

**Мета роботи.** Дослідити можливість застосування методів хемометрії для прогнозування ступеня небезпечності/ризиків залишкових розчинників у лікарських засобах за набором молекулярних дескрипторів.

**Матеріали і методи.** Об'єкт дослідження – класифікація залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах за ступенем ризику/небезпечності для здоров'я людини. Методи дослідження – тест Краскела – Уолліса; ймовірнісна нейронна мережа. Програмне забезпечення – програмний пакет ChemOffice 2020; програмний комплекс Matlab R2022b.

**Результати й обговорення.** Встановлено, що на класифікацію розчинників (за їх ступенем ризику/небезпечності відповідно до їх загрози здоров'ю людини) найбільший вплив мають такі молекулярні дескриптори (їх значення суттєво змінюються залежно від класу розчинника): кількість акцепторів водневого зв'язку; логарифмічний коефіцієнт розчинності у воді; площа полярної поверхні; коефіцієнт форми; сума валентних ступенів; загальна валентна зв'язність. Навчання ймовірнісної нейронної мережі на основі цих 6 молекулярних дескрипторів є коректним – з нульовою похибкою. Застосування ймовірнісної нейронної мережі забезпечує надійну класифікацію залишкових розчинників при широкому діапазоні значень відхилення функції активації.

**Висновки.** Запропоновано процедуру прогнозування ступеня небезпечності/ризиків залишкових розчинників у лікарських засобах.

**Вступ.** Органічні розчинники відіграють важливу роль у фармацевтичному виробництві, а саме при синтезі активних інгредієнтів і допоміжних речовин лікарських препаратів. Органічні розчинники можуть бути присутніми в добавках (консервантах, солубілізаторах, стабілізаторах) та в рецептурі лікарського засобу (в покритті таблетки або в матеріалі капсули, в розчинах настоянок, суспензій, емульсій, сиропів). Наявність органічних розчинників у невеликій кількості може вплинути на ефективність, безпеку та стабільність фармацевтичної продукції [1, 2].

Хоча органічні сполуки є леткими сполуками, але після завершення технологічного процесу видаляються не повністю. Особливістю органічних розчинників є їх токсичність незалежно від їх використання. У зв'язку з цим існує необхідність обов'язкового контролю їх вмісту в лікарських засобах.

Міжнародна конференція з гармонізації технічних вимог щодо реєстрації лікарських засобів для людини ухвалила Настанову з регламентації залишкових розчинників, що встановлює межі вмісту розчинників, які можуть залишатися в діючих речовинах, допоміж-

них речовинах і лікарських засобах на різних стадіях виробництва та у готових лікарських препаратах. Встановлено кількісні межі органічних розчинників і обґрунтовано дані про безпеку та токсичність розчинників, запропоновано межі і норми виробничих процесів у фармацевтичному виробництві [3].

За рівнем ризику органічні розчинники поділяються на три класи [3, 4]:

– розчинники 1 класу – найбільш токсичні та/або екологічно небезпечні (такі розчинники не повинні бути в ліках);

– розчинники 2 класу – середня категорія ризику (такі розчинники можуть бути в ліках в обмеженій кількості);

– розчинники 3 класу – найнижча категорія ризику (дозволена добова доза таких розчинників становить 50 мг/добу або більше).

На сьогодні є розчинники, що можуть становити інтерес для виробників діючих речовин, допоміжних речовин або лікарських засобів, але для них немає достатніх токсикологічних даних, якими можна обґрунтувати допустиму добову дозу. Тому, безумовно, є необхідність створення теоретичного (розрахункового) підходу щодо прогнозування ступеня небезпечності/ризиків залишкових розчинників у лікарських засобах. Метою даної роботи стало дослідження можливості застосування методів хемометрії для прогнозування ступеня небезпечності/ризиків залишкових розчинників у лікарських засобах за набором молекулярних дескрипторів. Під прогнозуванням ступеня ризику ми розуміємо приналежність розчинників до одного з трьох вказаних класів. Хемометричні методи, зокрема штучні нейронні мережі, успішно та ефективно використовуються у фармацевтичній промисловості від встановлення специфікацій контролю якості сировини, порошків і лікарських форм до контролю різних процесів і етапів виробництва [5].

**Матеріали і методи.** У даній роботі для дослідження можливості прогнозування ступеня ризику/небезпечності залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах були використані дані, що включені до Настанови з регламентації залишкових розчинників, ухвалену Міжнародною конференцією з гармонізації технічних вимог щодо реєстрації лікарських засобів для людини, та висвітлені у Державній Фармакопеї України [3, 4]. Масив даних містив:

– 59 розчинників, що за їхньою можливою загрозою здоров'ю людини класифіковані на три класи (таблиця 1): розчинники, використання яких слід уникати (клас 1); розчинники, використання яких слід обмежувати (клас 2); малотоксичні розчинники (клас 3);

– 9 розчинників<sup>1</sup>, що можуть становити інтерес для виробників лікарських засобів чи допоміжних речовин, для яких відсутні дані з токсичності (таблиця 2) [4].

Методи дослідження: тест Краскела – Уолліса для встановлення найбільш інформативних молекулярних дескрипторів для надійної класифікації залишкових розчинників за їх ступенем ризику/небезпечності [6]; ймовірнісна нейронна мережа для встановлення класової приналежності залишкових розчинників за їх ступенем ризику/небезпечності.

Ймовірнісна нейронна мережа – модифікація радіальних базисних нейронних мереж. Ймовірнісна нейронна мережа містить прихований шар нейронів з радіально-симетричною функцією активації, кожен із яких призначений для зберігання окремого еталонного зразка (відсутня потреба визначати число прихованих нейронів та вид функції активації):

$$F = e^{-\left[\frac{\sqrt{\sum_{i=1}^N (x_i - w_i)^2}}{2\delta}\right]^2}, \quad (1)$$

де  $\delta$  – відхилення функції активації,  $w_i$  – вагові коефіцієнти нейронів,  $x_i$  – числові характеристики досліджуваного об'єкта [7, 8].

Вихідний шар ймовірнісної мережі – конкуруючий шар, який підраховує ймовірність приналежності вхідного вектора до того чи іншого класу  $i$ , зрештою, зіставляє вектор з тим класом, ймовірність приналежності якого вище. Навчання ймовірнісної мережі передбачає попереднє проведення кластеризації для визначення центрів класів, найчастіше використовується алгоритм  $k$ -середніх [9, 10].

Основна проблема при реалізації ймовірнісної нейронної мережі полягає у визначенні параметра відхилення функції активації: значення параметра має бути досить великим, щоб перешкоджати перенавчанню, проте не настільки великим, щоб радіальна базова функція визначала однаково значущими всі значення входу. Вагові коефіцієнти першого шару формуються з використанням вхідних векторів з навчальної множини. Вагові коефіцієнти другого шару відповідають цільовим векторам навчальної вибірки [11].

**Результати й обговорення.**

#### **Розрахунок молекулярних дескрипторів**

За допомогою програмного пакета ChemOffice 2020 [12] для вказаних у таблицях 1 та 2 розчинників розраховано 18 молекулярних дескрипторів (хімічні, топологічні, фізичні та термодинамічні властивості) [13]:

- 1) молекулярна маса (molecular mass);
- 2) кількість акцепторів водневого зв'язку (number of HBond acceptors);
- 3) кількість донорів водневого зв'язку (number of HBond donors);

4) овальність (ovality):

$$O = A/(4 \times \pi \times ((3 \times V)/4 \times \pi)^{2/3}), \quad (2)$$

де  $O$  – овальність,  $A$  – площа,  $V$  – об'єм;

**Таблиця 1**

Розчинники, для яких відомі ступінь ризику/небезпеки відповідно до їх загрози здоров'ю людини

№ з/п	Назва розчинника	Клас розчинника
1	Анізол	3
2	Ацетон	3
3	Ацетонітрил	2
4	Бензол	1
5	1-Бутанол	3
6	2-Бутанол	3
7	Бутилацетат	3
8	<i>трет</i> -Бутилметиловий ефір	3
9	Гексан	2
10	Гептан	3
11	<i>N,N</i> -Диметилацетамід	2
12	Диметилсульфоксид	3
13	<i>N,N</i> -Диметилформамід	2
14	1,2-Диметоксіетан	2
15	1,4-Діоксан	2
16	Дихлорметан	2
17	1,2-Дихлоретан	1
18	1,1-Дихлоретен	1
19	1,2-Дихлоретен	2
20	Діетиловий ефір	3
21	Етанол	3
22	Етилацетат	3
23	Етиленгліколь	2
24	Етилформиат	3
25	2-Етоксіетанол	2
26	Ізобутилацетат	3
27	Ізопропілацетат	3
28	Ксилол	2
29	Кумол	2
30	Метанол	2

№ з/п	Назва розчинника	Клас розчинника
31	Метилацетат	3
32	3-Метил-1-бутанол	3
33	Метилбутилкетон	2
34	Метилізобутилкетон	3
35	2-Метил-1-пропанол	3
36	<i>N</i> -Метилпіролідон	2
37	Метилциклогексан	2
38	Метилетилкетон	3
39	2-Метоксіетанол	2
40	Мурашина кислота	3
41	Нітрометан	2
42	Пентан	3
43	1-Пентанол	3
44	Піридин	2
45	1-Пропанол	3
46	2-Пропанол	3
47	Пропілацетат	3
48	Сульфолан	2
49	Тетрагідрофуран	2
50	Тетралін	2
51	Толуол	2
52	1,1,1-Трихлоретан	1
53	1,1,2-Трихлоретен	2
54	Оцтова кислота	3
55	Формамід	2
56	Хлорбензол	2
57	Хлороформ	2
58	Циклогексан	2
59	Тетрахлорметан	1

<sup>1</sup>Петролейний ефір було виключено з переліку, оскільки це суміш насичених вуглеводнів, що унеможлиблює написання чіткої структурної формули та розрахунок молекулярних дескрипторів.

**Таблиця 2**

Розчинники, для яких відсутні достатні дані з токсичності

Назва розчинника	
1,1-Диметоксиметан	Метилізопропілкетон
2,2-Диметоксипропан	Метилтетрагідрофуран
1,1-Діетоксипропан	Трихлороцтова кислота
Діізопропіловий ефір	Трифтороцтова кислота
Ізооктан	

5) молярна рефракція (molar refractivity):

$$A = \frac{4\pi}{3} N_A \alpha, \quad (3)$$

де  $A$  – молярна рефракція,  $N_A$  – стала Авогадро,  $\alpha$  – середня поляризованість;

6) логарифмічний коефіцієнт розподілу між октанолом і водою ( $\log P$ ), що є показником ліпофільності;

7) логарифмічний коефіцієнт розчинності у воді ( $\log S$ );

8) індекс Балабана (Balaban index):

$$J = \frac{q}{\mu+1} \sum (S_i S_j)^{-1/2}, \quad (4)$$

де  $q$  – кількість ребер у молекулярному графі,  $\mu = q - n + 1$  – цикломатичне число молекулярного графа,  $n$  – число атомів в молекулярному графі,  $S_i$  – сума всіх елементів в  $i$ -му рядку (стовпчику) топологічної матриці відстаней;

9) кількість кластерів (cluster count) – кількість траєкторій заданої довжини в матриці відстаней;

10) молекулярний топологічний індекс (molecular topological index):

$$MTI = \sum_{i=1}^n E_i, \quad (5)$$

$$E = (A + D) * d, \quad (6)$$

де  $E$  – компоненти вектора,  $A$  – матриця суміжності,  $D$  – матриця відстані графа,  $d$  – вектор ступенів вершин графа;

11) площа полярної поверхні (polar surface area) – сума поверхні всіх полярних атомів (зазвичай Оксигену та Нітрогену), включно приєднані атоми Гідрогену;

12) атрибут форми (shape attribute) – вимірює розгалуження молекули;

13) коефіцієнт форми (shape coefficient):

$$I = \frac{D-R}{R}, \quad (7)$$

де  $D$  – максимальне значення діаметра між найбільш віддаленими атомами,  $R$  – мінімальне значення радіуса між найбільш центральними атомами;

14) сума ступенів (sum of degrees) – сума ступенів кожного атома (атомний ступінь – це кількість неводневих атомів, з якими він зв'язаний);

15) сума валентних ступенів (sum of valence degrees) – сума валентних ступенів кожного атома (валентний ступінь атома – це сума порядків його суцільних зв'язків, включно водні);

16) загальна зв'язність (total connectivity) – зв'язність, що розглядається для усіх гетероатомів;

17) загальна валентна зв'язність (total valence connectivity) – валентна зв'язність, що розглядається для всіх гетероатомів;

18) індекс Вінера (Wiener index):

$$W = \frac{1}{2} \sum D_{ij}, \quad (8)$$

де  $D_{ij}$  – недіагональні елементи матриці відстаней.

#### **Розрахунок критеріїв Краскела – Уолліса**

У даній роботі за допомогою критерію Краскела-Уолліса визначали молекулярні дескриптори, значення яких найбільше впливає на віднесення розчинника до того чи іншого класу. Розрахунок критерію Краскела – Уолліса для 59 розчинників (див. табл. 1), що характеризуються 18 молекулярними дескрипторами, виконано із застосуванням програмного комплексу Matlab R2022b [14] та наведено у таблиці 3. Якщо розрахункове значення критерію більше критичного, то тестований параметр суттєво змінюється залежно від класів розчинників, в інакшому випадку – немає статистично значимих міжкласових відмінностей для тестованого параметра [6].

Встановлено, що на класифікацію залишкових розчинників (за їх ступенем ризику/небезпеки відповідно до їх загрози здоров'ю людини) найбільший вплив становлять такі молекулярні дескриптори (їх значення суттєво змінюються залежно від класу розчинника): кількість акцепторів водневого зв'язку; логарифмічний коефіцієнт розчинності у воді; площа полярної поверхні; коефіцієнт форми.

Також суттєво на класифікацію залишкових розчинників впливають такі параметри, як сума валентних ступенів та загальна валентна зв'язність. Значення цих 6 молекулярних дескрипторів, що мають найбільший вплив на віднесення залишкових розчинників до певного класу, наведено у таблицях 4 та 5.

#### **Навчання ймовірнісної нейронної мережі**

Ймовірнісну нейронну мережу навчали на 59 розчинниках, для яких відомі ступінь ризику/небезпеки відповідно до їх загрози здоров'ю людини (див. табл. 1). Навчання проводили з використанням набору із 6 молекулярних дескрипторів (за результатами розрахунку критерію Краскела – Уоллі-

**Таблиця 3**

Результати розрахунку критерію Краскела – Уолліса для 18 молекулярних дескрипторів

	Номер молекулярного дескриптора																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
$\chi^2$ *	5,32	<b>11,06</b>	4,42	4,94	0,37	4,37	<b>6,23</b>	2,03	1,35	3,26	<b>10,3</b>	1,37	<b>9,24</b>	1,81	<b>5,75</b>	3,22	<b>5,92</b>	2,02

Примітка. \* – критичне значення  $\chi^2$  при рівні значущості 5 % та 10 %, відповідно, 5,99 та 4,60 (число ступенів свободи 2) [15].

**Таблиця 4**

Значення 6 молекулярних дескрипторів для розчинників із відомим ступенем ризику/небезпеки відповідно до їх загрози здоров'ю людини

Розчинник	Молекулярний дескриптор					
	кількість акцепторів водневого зв'язку	логарифмічний коефіцієнт розчинності у воді	площа полярної поверхні	коефіцієнт форми	сума валентних ступенів	загальна валентна зв'язність
1	2	3	4	5	6	7
Анізол	1	-1,6508	9,23	0	26,0000	0,0131
Ацетон	1	-0,1336	17,07	1	12,0000	0,2041
Ацетонітрил	1	-0,1284	23,79	1	10,0000	0,2236
Бензол	0	-1,7812	0,00	0	18,0000	0,0370
1-Бутанол	1	-0,3720	20,23	1	12,0000	0,1581
2-Бутанол	1	-0,3208	20,23	0	12,0000	0,1826
Бутилацетат	1	-1,1785	26,30	1	24,0000	0,0295
трет-Бутилметиловий ефір	1	-0,8691	9,23	0	14,0000	0,2041
Гексан	0	-2,7500	0,00	0	10,0000	0,2500
Гептан	0	-3,1641	0,00	1	12,0000	0,1768
N,N-Диметилацетамід	1	-0,0134	20,31	0	18,0000	0,0913
Диметилсульфоксид	1	0,3749	17,07	1	8,6667	0,5000
N,N-Диметилформамід	1	0,1983	20,31	0	16,0000	0,1054
1,2-Диметоксіетан	2	0,2558	18,46	0	18,0000	0,0833
1,4-Діоксан	2	0,0859	18,46	0	12,0000	0,0417
Дихлорметан	0	-1,4336	0,00	1	3,5556	0,9091
1,2-Дихлоретан	0	-1,5948	0,00	0	5,5556	0,6429
1,1-Дихлоретен	0	-1,9385	0,00	1	7,5556	0,4546
1,2-Дихлоретен	0	-2,1835	0,00	0	7,5556	0,4286
Діетиловий ефір	1	-0,5100	9,23	1	12,0000	0,2041
Етанол	1	0,3046	20,23	1	8,0000	0,3162
Етилацетат	1	-0,5017	26,30	1	20,0000	0,0589
Етиленгліколь	2	1,0049	40,46	0	14,0000	0,1000
Етилформиат	1	-0,2904	26,30	1	18,0000	0,0680
2-Етоксіетанол	2	0,4137	29,46	0	18,0000	0,0645
Ізобутилацетат	1	-1,0107	26,30	0	24,0000	0,0340
Ізопропілацетат	1	-0,8665	26,30	1	22,0000	0,0481
Ксилол	0	-2,5711	0,00	0	22,0000	0,0278
Кумол	0	-2,9726	0,00	0	24,0000	0,0185
Метанол	1	0,6361	20,23	0	6,0000	0,4472
Метилацетат	1	-0,1680	26,30	0	18,0000	0,0833
3-Метил-1-бутанол	1	-0,7697	20,23	1	14,0000	0,1291
Метилбутилкетон	1	-1,1641	17,07	0	18,0000	0,0722
Метилізобутилкетон	1	-1,1465	17,07	1	18,0000	0,0833
2-Метил-1-пропанол	1	-0,2042	20,23	0	12,0000	0,1826
N-Метилпіролідон	1	-0,2788	20,31	0	22,0000	0,0323

1	2	3	4	5	6	7
Метилциклогексан	0	-3,2048	0,00	0	14,0000	0,1021
Метилетилкетон	1	-0,3374	17,07	0	14,0000	0,1443
2-Метоксіетанол	2	0,7457	29,46	1	16,0000	0,0913
Мурашина кислота	1	0,4091	37,30	1	14,0000	0,1054
Нітрометан	1	-0,4563	51,81	1	18,0000	0,0745
Пентан	0	-2,3358	0,00	1	8,0000	0,3536
1-Пентанол	1	-0,7861	20,23	0	14,0000	0,1118
Піридин	1	-0,8595	12,36	0	20,0000	0,0287
1-Пропанол	1	0,0420	20,23	0	10,0000	0,2236
2-Пропанол	1	-0,0580	20,23	1	10,0000	0,2582
Пропілацетат	1	-0,7644	26,30	0	22,0000	0,0417
Сульфолан	2	-1,2008	34,14	0	20,6667	0,0510
Тетрагідрофуран	1	-0,5119	9,23	0	14,0000	0,1021
Тетралін	0	-3,2714	0,00	0	28,0000	0,0069
Толуол	0	-2,1816	0,00	0	20,0000	0,0321
1,1,1-Трихлоретан	0	-2,3969	0,00	1	7,3333	0,7289
1,1,2-Трихлоретен	0	-2,6907	0,00	0	9,3333	0,4208
Оцтова кислота	1	0,1980	37,30	1	16,0000	0,0913
Формамід	1	0,5815	43,09	1	12,0000	0,1361
Хлорбензол	0	-2,5203	0,00	0	19,7778	0,0364
Хлороформ	0	-2,0560	0,00	1	5,3333	0,8417
Циклогексан	0	-2,8068	0,00	0	12,0000	0,1250
Тетрахлорметан	0	-2,8769	0,00	1	7,1111	0,8265

**Таблиця 5**

Значення 6 молекулярних дескрипторів для розчинників із невідомим ступенем ризику/небезпеки відповідно до їх загрози здоров'ю людини

Розчинник	Молекулярний дескриптор					
	кількість акцепторів водневого зв'язку	логарифмічний коефіцієнт розчинності у воді	площа полярної поверхні	коефіцієнт форми	сума валентних ступенів	загальна валентна зв'язність
1,1-Диметоксиметан	2	0,1822	18,46	1	16	0,1179
2,2-Диметоксипропан	2	-0,5374	18,46	1	20	0,0833
1,1-Діетоксипропан	2	-1,2031	18,46	1	24	0,0417
Діізопропіловий ефір	1	-1,2361	9,23	1	16	0,1361
Ізооктан	0	-3,5932	0,00	1	14	0,2041
Метилізопропілкетон	1	-1,1465	17,07	1	18	0,0833
Метилтетрагідрофуран	1	-0,8752	9,23	0	16	0,0833
Трихлороцтова кислота	1	-1,1525	37,3	0	21,3333	0,0665
Трифтороцтова кислота	4	-0,2425	37,3	0	40	0,0025



са) та при різних значеннях параметру відхилення функції активації. Для контролю якості навчання ймовірнісної нейронної мережі 59 розчинників поділили випадковим чином на навчальну вибірку (80 %), тестову вибірку (10 %) та контрольну (валідаційну) вибірку (10 %). Склад тестової та контрольної вибірок наведено у таблиці 6; у навчальну вибірку увійшли усі розчинники, що не потрапили до цих вибірок.

Ненадійність класифікації оцінювали як частку невірно класифікованих зразків тестової (контрольної) вибірки [16]:

$$P = \frac{n}{N} * 100 \%, \quad (9)$$

де,  $n$  – число невірно класифікованих зразків тестової (контрольної) вибірки,  $N$  – загальне число зразків тестової (контрольної) вибірки.

У таблиці 7 наведено результати оцінки навчання ймовірнісної нейронної мережі з використанням 6 молекулярних дескрипторів, відбір яких базувався на результатах розрахунку критерію Краскела – Уолліса (при  $\delta=0,1$  спостерігається невірне віднесення до 1 класу залишкового розчинника контрольної вибірки N-Метилпіролідону).

**Прогнозування ступеня ризику/небезпеки залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах**

Для прогнозування ступеня ризику/небезпеки залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах використали ймовірнісну штучну нейронну мережу навчену на 59 розчинниках, що характеризуються 6 молекулярними дескрипторами: кількість акцепторів водневого зв'язку, логарифмічний коефіцієнт розчинності у воді, площа полярної поверхні, коефіцієнт форми, сума валентних ступенів, загальна валентна зв'язність.

Залишкові розчинники, для яких прогнозували ступінь ризику/небезпеки відповідно до їх загрози здоров'ю, представлено у таблиці 2. Результати прогнозування ступеня ризику/небезпеки залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах для 9 розчинників представлені у таблиці 8. Спостерігається однакова класифікація 9 розглянутих залишкових розчинників при значеннях відхилення функції активації від 0,1 до 1,0, крім трихлороцтової кислоти: при  $\delta=0,1$  вона класифікована як залишковий розчинник 1 класу, при  $\delta = 0,2-1,0$  – як залишковий розчинник 2 класу.

**Таблиця 6**

Тестова та контрольна вибірки для контролю якості навчання ймовірнісної нейронної мережі

Тестова вибірка	Контрольна вибірка
1,2-Диметоксіетан	N,N-Диметилацетамід тетрахлорметан
1-Пентанол	2-Метил-1-пропанол
1-Пропанол	N-Метилпіролідон
2-Пропанол	3-Метил-1-бутанол
1-Бутанол	Мурашина кислота
Гексан	

**Таблиця 7**

Результати навчання ймовірнісної нейронної мережі з використанням 6 молекулярних дескрипторів (кількість акцепторів водневого зв'язку, логарифмічний коефіцієнт розчинності у воді, площа полярної поверхні, коефіцієнт форми, сума валентних ступенів, загальна валентна зв'язність)

Параметр $\delta$	P, %	
	тестова вибірка	контрольна вибірка
0,1	0,0	<b>16,7</b>
0,2	0,0	0,0
0,3	0,0	0,0
0,4	0,0	0,0
0,5	0,0	0,0
0,6	0,0	0,0
0,7	0,0	0,0
0,8	0,0	0,0
0,9	0,0	0,0
1,0	0,0	0,0

Таблиця 8

Результати прогнозування ступеня ризику/небезпеки залишкових розчинників у субстанціях, допоміжних речовинах і лікарських засобах для 9 розчинників

Розчинник	Номер класу	
	$\delta = 0,1$	$\delta = 0,2-1,0$
1,1-Диметоксиметан	2	2
2,2-Диметоксипропан	2	2
1,1-Діетоксипропан	2	2
Діізопропіловий ефір	3	3
Ізооктан	2	2
Метилізопропілкетон	3	3
Метилтетрагідрофуран	3	3
Трихлороцтова кислота	1	2
Трифтороцтова кислота	1	1

**Висновки.** Набір із шести молекулярних дескрипторів (кількість акцепторів водневого зв'язку, логарифмічний коефіцієнт розчинності у воді, площа полярної поверхні, коефіцієнт форми, сума валентних ступенів, загальна валентна зв'язність) є достатнім та оптимальним для правильної класифікації залишкових розчинників за їх ступенем ризику/небезпеки.

Застосування ймовірнісної нейронної мережі забезпечує надійну класифікацію залишкових розчинників за їх можливою загрозою здоров'ю людини при

широкому діапазоні значень відхилення функції активації.

Запропонована процедура прогнозування ступеня ризику/небезпеки залишкових розчинників у лікарських засобах може використовуватися при плануванні та створенні нових синтетичних лікарських засобів.

**Конфлікт інтересів:** відсутній.

**Conflicts of interest:** authors have no conflict of interest to declare.

## PREDICTION OF THE DEGREE OF DANGER/RISK OF RESIDUAL SOLVENTS IN MEDICINAL PRODUCTS USING CHEMOMETRIC METHODS

Y. M. Pushkarova, A. V. Kaliuzhenko

Bogomolets National Medical University  
yaroslava.pushkarova@gmail.com

**The aim of the work.** To investigate the applicability of chemometric methods for prediction of risk assessment for residual solvents in pharmaceuticals based on a set of molecular descriptors.

**Materials and Methods.** The object of the study is the classification of residual solvents in drug substances, excipients, and in drug products according to the degree of risk/danger to human health. Research methods – Kruskal-Wallis test; probabilistic neural network. Software – software packages MATLAB R2022b and ChemOffice 2020.

**Results and Discussion.** It was established that the classification of solvents according to their risk to human health is most influenced by the following molecular descriptors (their values change significantly depending on the class of solvents): number of HBond acceptors, logarithm of the solubility of a substance, polar surface area, shape coefficient, sum of valence degrees, total valence connectivity. The correct training of probabilistic neural network was archived in the case of using above six molecular descriptors at different values of spread parameter of the Gaussian activation function (there are no classification errors for testing and validation subsets).

**Conclusions.** Probabilistic neural network is the effective tool for risk assessment of residual solvents.

**Key words:** residual solvent; classification; descriptor; artificial neural network.



**Перелік бібліографічних посилань**

1. Q-SAVESS: a methodology to help solvent selection for pharmaceutical manufacture at the early process development stage / V. Isoni et al. *Green Chemistry*. 2016. Vol. 18, no. 24. P. 6564–6572. URL: <https://doi.org/10.1039/c6gc02440h> (date of access: 18.07.2023).
2. Papadakis E., Tula A. K., Gani R. Solvent selection methodology for pharmaceutical processes: Solvent swap. *Chemical Engineering Research and Design*. 2016. Vol. 115. P. 443–461. URL: <https://doi.org/10.1016/j.cherd.2016.09.004> (date of access: 18.07.2023).
3. Impurities: Guideline for residual solvents. European Medicines Agency, Amsterdam, Netherlands, 2021, 51 p. URL: <https://www.ich.org/page/quality-guidelines> (date of access: 18.07.2023).
4. Державна Фармакопея України / Державне підприємство «Український науковий фармакопейний центр якості лікарських засобів». – 2-е вид. – Доповнення 2. Харків : Державне підприємство «Український науковий фармакопейний центр якості лікарських засобів», 2018. – 5.4. Залишкові розчинники. URL: <http://sphu.org/viddil-dfu> (дата звернення: 18.07.2023).
5. Kaliuzhenko A., Pushkarova Y. Application of artificial neural networks for solving pharmaceutical issues. *Grail of Science*. 2023. No. 24. P. 766–769. URL: <https://doi.org/10.36074/grail-of-science.17.02.2023.143> (date of access: 18.07.2023).
6. Ostertagová E., Ostertag O., Kováč J. Methodology and Application of the Kruskal-Wallis Test. *Applied Mechanics and Materials*. 2014. Vol. 611. P. 115–120. URL: <https://doi.org/10.4028/www.scientific.net/amm.611.115> (date of access: 18.07.2023).
7. Zeinali Y., Story B. A. Competitive probabilistic neural network. *Integrated Computer-Aided Engineering*. 2017. Vol. 24, No. 2. P. 105–118. URL: <https://doi.org/10.3233/ica-170540> (date of access: 18.07.2023).
8. Savchenko A. V., Belova N. S. Sequential analysis in fourier probabilistic neural networks. *Expert Systems with Applications*. 2022. P. 117885. URL: <https://doi.org/10.1016/j.eswa.2022.117885> (date of access: 18.07.2023).
9. Hoya T. Reducing the number of centers in a probabilistic neural network via applying the first neighbor means clustering algorithm. *Array*. 2022. Vol. 14. P. 100161. URL: <https://doi.org/10.1016/j.array.2022.100161> (date of access: 18.07.2023).
10. Ahmed M., Seraj R., Islam S. M. S. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*. 2020. Vol. 9, No. 8. P. 1295. URL: <https://doi.org/10.3390/electronics9081295> (date of access: 18.07.2023).
11. Pushkarova Y., Kholin Y. A procedure for meaningful unsupervised clustering and its application for solvent classification. *Open Chemistry*. 2014. Vol. 12, No. 5. P. 594–603. URL: <https://doi.org/10.2478/s11532-014-0514-6> (date of access: 18.07.2023).
12. The Prime Chemistry Portal. URL: <https://chemistry-docs.com/perkinelmer-chemoffice-2020-version-20-0/> (date of access: 18.07.2023).
13. Columbia University Libraries, Chem3D 17.0 User Guide. Copyright 1998-2017 PerkinElmer Informatics Inc., p. 203-232. URL: [https://library.columbia.edu/content/dam/libraryweb/locations/dsc/Software%20Subpages/ChemDraw\\_17\\_manual.pdf](https://library.columbia.edu/content/dam/libraryweb/locations/dsc/Software%20Subpages/ChemDraw_17_manual.pdf) (date of access: 18.07.2023).
14. Matlab for artificial intelligence. URL: <https://www.mathworks.com/products/matlab.html> (date of access: 18.07.2023).
15. Miller R. D., Miller J. C., Miller J. Statistics and Chemometrics for Analytical Chemistry. Pearson Education Australia, 2018. 296 p.
16. Pushkarova Y., Kholin Y. The classification of solvents based on solvatochromic characteristics: the choice of optimal parameters for artificial neural networks. *Open Chemistry*. 2012. Vol. 10, No. 4. P. 1318–1327. URL: <https://doi.org/10.2478/s11532-012-0060-z> (date of access: 18.07.2023).

**References**

1. Isoni V, Wong LL, Khoo HH, Halim I, Sharratt P. Q-SAVESS: a methodology to help solvent selection for pharmaceutical manufacture at the early process development stage. *Green Chem*. 2016;18(24): 6564-72. <https://doi.org/10.1039/c6gc02440h>
2. Papadakis E, Tula AK, Gani R. Solvent selection methodology for pharmaceutical processes: Solvent swap. *Chem Eng Res Des*. 2016;115: 443-61. <https://doi.org/10.1016/j.cherd.2016.09.004>
3. Impurities: Guideline for residual solvents. Amsterdam, Netherlands: European Medicines Agency; 2021. <https://www.ich.org/page/quality-guidelines>
4. The State Pharmacopoeia of Ukraine. 2<sup>nd</sup> ed. Residual Solvents. Ukraine, Kharkiv: Ukrainian Scientific Pharmacopoeial Center for Quality of Medicines; 2018. 12 p. Ukrainian. <http://sphu.org/viddil-dfu>
5. Kaliuzhenko A, Pushkarova Y. Application of artificial neural networks for solving pharmaceutical issues. *Grail Sci*. 2023;(24): 766-9. <https://doi.org/10.36074/grail-of-science.17.02.2023.143>
6. Ostertagová E, Ostertag O, Kováč J. Methodology and Application of the Kruskal-Wallis Test. *Appl Mech Mater*. 2014;611: 115-20. <https://doi.org/10.4028/www.scientific.net/amm.611.115>
7. Zeinali Y, Story BA. Competitive probabilistic neural network. *Integr Comput Aided Eng*. 2017;24(2): 105-18. <https://doi.org/10.3233/ica-170540>
8. Savchenko AV, Belova NS. Sequential analysis in fourier probabilistic neural networks. *Expert Syst With Appl*. 2022: 117885. <https://doi.org/10.1016/j.eswa.2022.117885>
9. Hoya T. Reducing the number of centers in a probabilistic neural network via applying the first neighbor means clustering algorithm. *Array*. 2022;14: 100161.

- <https://doi.org/10.1016/j.array.2022.100161>
- Ahmed M, Seraj R, Islam SM. The k-means Algorithm: A Comprehensive Survey and Performance Evaluation. *Electronics*. 2020;9(8): 1295. <https://doi.org/10.3390/electronics9081295>
  - Pushkarova Y, Kholin Y. A procedure for meaningful unsupervised clustering and its application for solvent classification. *Open Chem*. 2014;12(5): 594-603. <https://doi.org/10.2478/s11532-014-0514-6>
  - The Prime Chemistry Portal. URL: <https://chemistry-docs.com/perkinelmer-chemoffice-2020-version-20-0/>
  - Columbia University Libraries, Chem3D 17.0 User Guide. Copyright 1998-2017 PerkinElmer Informatics Inc., URL: [https://library.columbia.edu/content/dam/libraryweb/locations/dsc/Software%20Subpages/ChemDraw\\_17\\_manual.pdf](https://library.columbia.edu/content/dam/libraryweb/locations/dsc/Software%20Subpages/ChemDraw_17_manual.pdf)
  - Matlab for artificial intelligence. URL: <https://www.mathworks.com/products/matlab.html>
  - Miller RD, Miller JC, Miller J'. *Statistics and Chemometrics for Analytical Chemistry*. Australia: Pearson Education; 2018. 296 p.
  - Pushkarova Y, Kholin Y. The classification of solvents based on solvatochromic characteristics: the choice of optimal parameters for artificial neural networks. *Open Chem*. 2012;10(4): 1318-27. <https://doi.org/10.2478/s11532-012-0060-z>

#### **Відомості про авторів**

**Пушкарьова Я. М.** – канд. хім. наук, доцент закладу вищої освіти кафедри аналітичної, фізичної та колоїдної хімії, Національний медичний університет імені О. О. Богомольця, Київ, Україна. E-mail: [yaroslava.pushkarova@gmail.com](mailto:yaroslava.pushkarova@gmail.com), ORCID 0000-0001-9856-7846.

**Калюженко А. В.** – магістр (фармація), Київ, Україна. E-mail: [0992094671a@gmail.com](mailto:0992094671a@gmail.com)

**Pushkarova Ya. M.** – PhD (Chemistry), Associate Professor of the Analytical, Physical and Colloid Chemistry Department, Bogomolets National Medical University, Kyiv, Ukraine. E-mail: [yaroslava.pushkarova@gmail.com](mailto:yaroslava.pushkarova@gmail.com), ORCID 0000-0001-9856-7846.

**Kaliuzhenko A. V.** – master (Pharmacy), Kyiv, Ukraine. E-mail: [0992094671a@gmail.com](mailto:0992094671a@gmail.com)