UDC 543.33 + 004.032.26
DOI: 10.20535/SRIT.2308-8893.2025.4.07

PREDICTION OF MECHANISMS OF TOXIC ACTION OF PHENOLS BY MEANS OF PROBABILISTIC NEURAL NETWORK IN COMBINATION WITH KRUSKAL–WALLIS TEST

Ya.M. PUSHKAROVA, G.M. ZAITSEVA

Abstract. Prediction of the toxicity of chemical compounds is one of the most important steps in drug design. The use of phenolic compounds is a promising component in the pharmaceutical industry with many possible applications. The paper focuses on the application of a probabilistic neural network for classifying 232 phenols based on their mechanisms of toxic action. The Kruskal–Wallis test was also used to assess the influence of molecular descriptors on the reliable classification of phenolic compounds based on the mechanisms of their toxic action. It is shown that for the correct training of a probabilistic neural network and effective prediction of the mechanisms of toxic action of phenols, it is sufficient to use only 5 molecular descriptors.

Keywords: artificial neural network, classification, drug design, phenol, toxicity.

INTRODUCTION

Assessment of the toxicity of chemical compounds is an important and necessary stage on the way to the creation of new medicines. It is known that the experimental study of only one type of toxicity is an expensive and long-term process. Phenolic compounds have a number of useful properties that make them interesting for pharmacy: antioxidant, anti-inflammatory, antimicrobial properties, anticancer activities, etc. Additionally, phenolic compounds are often found in natural sources, such as plants, which adds to their appeal for use in pharmacy [1–4].

Overall, the diverse range of beneficial properties exhibited by phenolic compounds makes them valuable compounds in pharmacy and medicine, with potential applications in the treatment and prevention of various diseases. But before using phenols in pharmacy, it is important to predict possible mechanisms of their toxic action (polar narcotics, weak acid respiratory uncouplers, pro-electrophiles and soft electrophiles). This helps to identify risks to people and to take measures to reduce the possible negative consequences, that is, to develop safe medicines [5; 6].

Chemometric methods use mathematical and statistical models to analyze complex data sets and extract meaningful information, making them valuable tools in pharmaceutical research and development. Chemometric methods, in particular artificial neural networks, are widely used for prediction and classification tasks in pharmacy. Artificial neural networks are computational models inspired by the structure and functioning of biological neural networks in the human brain. These methods can help predict various properties of pharmaceutical compounds, such as their stability, toxicity, solubility and bioavailability. They are also used for identifying different types of drugs or distinguishing between counterfeit and authentic products [7–10].

MATERIALS AND METHODS

Data Set

The studied dataset consists of a training, testing and validation sub-sets with a total of 232 phenolic compounds: training sub-set – 197 phenols, testing sub-set – 20 phenols, validation sub-set – 15 phenols. All phenolic compounds were characterized by seven physical-chemical descriptors: 1) distribution coefficient; 2) energy of the lowest unoccupied molecular orbital; 3) molecular weight; 4) negatively charged molecular surface area in percent's; 5) sum of absolute charges on nitrogen and oxygen atoms in a molecule; 6) largest positive charge on a hydrogen atom; 7) electrotopological state index for the hydroxyl group. Values of these descriptors and toxicity values were taken from [6].

Distribution of the studied phenolic compounds into classes according to the mechanisms of toxic action of phenolic compounds to Tetrahymena pyriformis is presented in Table 1. The most numerous class is class 1 of polar narcotics (71.6% of all studied phenolic compounds), other classes are almost the same in number of samples.

Table 1. Distribution of the studied phenolic compounds into classes according to the mechanisms of toxic action to Tetrahymena pyriformis

Classes According to Mechanisms of Toxic Action	Number of Phenolic Compounds			
	Training sub-set	Testing sub-set	Validation sub-set	Total
Class 1. Polar narcotics	138	16	12	166
Class 2. Weak acid respiratory uncouplers	15	1	1	17
Class 3. Pro-electrophiles	22	2	0	24
Class 4. Soft electrophiles	22	1	2	25

Applied Methods

The software package Matlab R2023b (trial individual license 11937601) was used in the present work for realization Kruskal–Wallis test and probabilistic neural network [11].

The Kruskal–Wallis test is a non-parametric statistical test used to determine whether there are statistically significant differences between two or more groups of a dependent variable [12].

A probabilistic neural network is a type of artificial neural network, which consists of following layers: input layer, pattern layer, summation layer, and output layer. A brief overview of how probabilistic neural network works [13–15]:

- input layer receives the input pattern;
- neurons of pattern layer store the training patterns;
- summation layer computes the similarity between the input pattern and the stored patterns using Gaussian function;
- output layer produces the class probability estimates.

To classify a new input pattern, the probabilistic neural network computes the class probabilities using the summation layer and outputs the class with the highest probability.

RESULTS AND DISCUSSION

Definition of Informative Descriptors for Classification of Phenolic Compounds into Classes According to the Mechanisms of Toxic Action

The calculation of the Kruskal–Wallis test for 232 phenols characterized by 7 molecular descriptors and toxicity is given in Table 2.

Table 2. Results of the Kruskal–Wallis test calculation for 7 descriptors and toxicity

Parameter	Toxicity	Distribution coefficient	Energy of the lowest unoccupied molecular orbita	Molecular weight	Negatively charged molecular surface area in percent's	Sum of absolute charges on nitrogen and oxygen atoms in a molecule	Largest positive charge on a hydrogen atom	Electrotopological state index for the hydroxyl group
χ^2	17.80	54.32	104.90	35.78	70.24	31.71	4.34	18.56

Critical value of χ^2 at the significance level of 5% with 3 degrees of freedom is 7.82 [16].

It was established some dependences between studied descriptors and classification of phenolic compounds according to the mechanisms of their toxic action:

1) descriptor largest positive charge on a hydrogen atom is not influenced on classification of phenolic compounds according to the mechanisms of toxic action, because experimental value of $\chi 2$ is less than critical value ($4.34 < 7.82$);

2) descriptor energy of the lowest unoccupied molecular orbital has the greatest influence on the phenols classification according to the mechanisms of toxic action (maximum experimental value of χ^2 is established for this descriptor — 104.90);

3) the studied parameters can be conventionally divided into three groups according to their influence on the classification of phenols:

- weak influence: toxicity and electrotopological state index for the hydroxyl group;
- moderately strong influence: molecular weight and sum of absolute charges on nitrogen and oxygen atoms in a molecule;
- strong influence: distribution coefficient, energy of the lowest unoccupied molecular orbital and negatively charged molecular surface area in percent's.

Application of Probabilistic Neural Network

In the context of the probabilistic neural network, the spread of the Gaussian function is an important parameter for its construction. Choosing the right spread parameter is crucial for the performance of the probabilistic neural network. If the spread is too small, the network may over fit to the training data and perform poorly on new data. If the spread is too large, the network may under fit and fail to capture the underlying patterns in the data [8; 13].

In the present work it was investigated the applicability of probabilistic neural network at different values of the spread of the Gaussian function: 0.1; 0.2;

0.3; 0.4; 0.5; 0.6; 0.7; 0.8; 0.9; 1.0. It should be noted that the probabilistic neural network is trained with zero error at spread values from 0.1 to 1.0 for different sets of descriptors. Results of prediction of the mechanisms of toxic action of phenols for testing and validation sub-sets are also the same for spread values from 0.1 to 1.0 for different sets of descriptors.

The unreliability of the prediction was estimated as the part of incorrectly classified phenols of the testing or validation sub-sets in percent's [8]:

$$P = \frac{n}{N} \cdot 100\% ,$$

where n is the number of incorrectly classified phenols in the testing or validation sub-set; N is the total number of phenols in the testing or validation sub-set.

Results of prediction of the mechanisms of toxic action of phenolic compounds by means of probabilistic neural network based on a set of all 7 molecular descriptors and toxicity are shown in Table 3.

Table 3. Unreliability values of the prediction based on a set of all 7 molecular descriptors and toxicity

Sub-set	***P*, %**
Testing	10.0
Validation	6.7

Results of prediction of the mechanisms of toxic action of phenolic compounds by means of probabilistic neural network based on a set of 5 molecular descriptors (distribution coefficient, energy of the lowest unoccupied molecular orbital, molecular weight, negatively charged molecular surface area in percent's and sum of absolute charges on nitrogen and oxygen atoms in a molecule) are shown in Table 4.

Table 4. Unreliability values of the prediction based on a set of 5 molecular descriptors

Sub-set	***P*, %**
Testing	20.0
Validation	6.7

One can see, that results of prediction of the mechanisms of toxic action of phenolic compounds based on a set of all 7 molecular descriptors with toxicity and based on a set of 5 molecular descriptors are differed by two incorrectly classified phenols. This confirms, the verity of calculation results of the Kruskal–Wallis test: largest positive charge on a hydrogen atom, toxicity and electrotopological state index for the hydroxyl group are weakly influenced on assignment of phenols to one or another class according to mechanisms of their toxic action.

Decreasing the number of descriptors into 3 (distribution coefficient, energy of the lowest unoccupied molecular orbital and negatively charged molecular surface area in percent's) resulted in an increasing the part of incorrectly classified phenols of the testing sub-set from 20% till 40% (Table 5). It means, that molecular weight and sum of absolute charges on nitrogen and oxygen atoms in a mole-

cule are moderately strong influenced for classification of phenols according to mechanisms of their toxic action and can't be ignore.

Table 5. Unreliability values of the prediction based on a set of 3 molecular descriptors

Sub-set	*P*, %
Testing	40.0
Validation	6.7

Detailed information about prediction of the mechanisms of toxic action of phenolic compounds of testing and validation sub-sets are shown in Tables 6 and 7, correspondingly: 1 — polar narcotics; 2 — weak acid respiratory uncouplers; 3 — pro-electrophiles; 4 — soft electrophiles. Incorrect predictions are indicated in bold text.

Table 6. Results of prediction of the mechanisms of toxic action of phenols of the testing sub-set

N	Phenol compound	Predicted mechanism of toxic action using 7 descriptors and toxicity (0.1 ≤ spread ≤ 1.0)	Predicted mechanism of toxic action using 5 descriptors (0.1 ≤ spread ≤ 1.0)	Predicted mechanism of toxic action using 3 descriptors (0.1 ≤ spread ≤ 1.0)	True mechanism of toxic action [5, 6]
1	2-Fluorophenol	1	1	1	1
2	2-Allylphenol	1	1	1	1
3	3-Chlorophenol	1	1	1	1
4	4,6-Dichlororesorcinol	1	1	**3**	1
5	4-Benzyloxyphenol	1	1	1	1
6	3-Iodophenol	1	1	1	1
7	2,3-Dichlorophenol	1	1	1	1
8	4-Phenylphenol	1	1	1	1
9	4-Hexyloxyphenol	1	1	**3**	1
10	4-Hexylresorcinol	1	1	1	1
11	2,4,5-Trichlorophenol	1	1	1	1
12	2,4-Diaminophenol	3	3	**1**	3
13	Methylhydroquinone	3	**1**	**1**	3
14	3-Nitrophenol	4	4	**1**	4
15	4-Ethoxyphenol	1	**3**	**3**	1
16	4-Bromo-2,6-dimethylphenol	1	1	1	1
17	4-Methoxyphenol	1	1	1	1
18	2,6-Diiodo-4-nitrophenol	**1**	**1**	**4**	2
19	2-Methyl-3-nitrophenol	**4**	**4**	**4**	1
20	4-Isopropylphenol	1	1	1	1

Table 7. Results of prediction of the mechanisms of toxic action of phenols of the validation sub-set

N	Phenol compound	Predicted mechanism of toxic action using 7 descriptors and toxicity (0.1 ≤ spread ≤ 1.0)	Predicted mechanism of toxic action using 5 descriptors (0.1 ≤ spread ≤ 1.0)	Predicted mechanism of toxic action using 3 descriptors (0.1 ≤ spread ≤ 1.0)	True mechanism of toxic action [5, 6]
1	4-Hydroxypropiophenone	1	1	1	1
2	3-Hydroxybenzaldehyde	1	1	1	1
3	4-(4-Hydroxyphenyl)-2-butanone	1	1	1	1
4	4-Hydroxybenzaldehyde	1	1	1	1
5	4-Isopropylphenol	1	1	1	1
6	3-Fluoro-4-nitrophenol	4	4	4	4
7	Benzyl-4-hydroxybenzoate	1	1	1	1
8	5-Pentylresorcinol	1	1	1	1
9	2-Hydroxy-4-methoxyacetophenone	1	1	1	1
10	3-Methyl-2-nitrophenol	1	1	1	1
11	2-Ethylhexyl-4′-hydroxybenzoate	1	1	1	1
12	2,3-Dinitrophenol	2	2	**1**	2
13	2-Nitrophenol	4	4	4	4
14	3-Methoxyphenol	1	1	1	1
15	4-Chlororesorcinol	**3**	**3**	1	1

CONCLUSIONS

A set of five molecular descriptors (distribution coefficient, energy of the lowest unoccupied molecular orbital, molecular weight, negatively charged molecular surface area in percent's and sum of absolute charges on nitrogen and oxygen atoms in a molecule) is sufficient for correct classification of phenolic compounds by mechanisms their toxic effects.

The application of probabilistic neural network provides a reliable classification of phenolic compounds by mechanisms of their toxic action, as well as prediction of the mechanisms of their toxic action with high accuracy.

The proposed procedure for predicting the mechanisms of toxic action of phenolic compounds can be useful at the stage of development of medicines.

REFERENCES

1. R.E. Mutha, A.U. Tatiya, S.J. Surana, "Flavonoids as natural phenolic compounds and their role in therapeutics: An overview," *FJPS*, vol. 7, article no. 25, pp. 1–13, 2021. doi: https://doi.org/10.1186/s43094-020-00161-8
2. A. Durazzo et al., "Polyphenols: A concise overview on the chemistry, occurrence, and human health," *Phytother. Res.*, vol. 33(9), pp. 2221–2243, 2019. doi: https://doi.org/10.1002/ptr.6419
3. M.M. Rahman et al., "Role of phenolic compounds in human disease: current knowledge and future prospects," *Molecules*, vol. 27(1), pp. 233, 2021. doi: https://doi.org/10.3390/molecules27010233
4. Y. Pushkarova, G. Zaitseva, M.Al Saker, "Prediction of Toxicity of Phenols Using Artificial Neural Networks," *IEEE — 12th International Conference on Advanced Computer Information Technologies (ACIT), Spisska Kapitula, Slovakia, 26–28 September 2022*, pp. 493–496. doi: https://doi.org/10.1109/ACIT54803.2022.9913174

5. N. Aptula et al., "Multivariate discrimination between modes of toxic action of phenols," *QSAR*, vol. 21(1), pp. 12–22, 2002. doi: https://doi.org/10.1002/1521-3838(200205)21:1<12::AID-QSAR12>3.0.CO;2-M
6. M.T. Cronin et al., "Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to Tetrahymena pyriformis," *Chemosphere*, vol. 49(10), pp. 1201–1221, 2002. doi: https://doi.org/10.1016/s0045-6535(02)00508-8
7. Y. Pushkarova, V. Panchenko, Y. Kholin, "Application an Artificial Neural Network for Prediction of Substances Solubility," *IEEE EUROCON 2021 − 19th International Conference on Smart Technologies, Lviv, Ukraine, 6–8 July 2021*, pp. 82–87. doi: https://doi.org/10.1109/EUROCON52738.2021.9535593
8. Y. Pushkarova, G. Zaitseva, A. Kaliuzhenko, "Classification of Residual Solvents by Risk Assessment Using Chemometric Methods," *IEEE — 13 th International Conference on Advanced Computer Information Technologies* (*ACIT*)*, Wroclaw, Poland, 21–22 September 2023*, pp. 562–565. doi: https://doi.org/10.1109/ACIT58437. 2023.10275405
9. E. Havránková, E.M. Pena-Mendez, J. Csöllei, J. Havel, "Prediction of biological activity of compounds containing a 1, 3, 5-triazinyl sulfonamide scaffold by artificial neural networks using simple molecular descriptors," *Bioorg. Chem.*, vol. 107, pp. 104565, 2021. doi: https://doi.org/10.1016/j.bioorg.2020.104565
10. S. Gummadi, P.K. Chandaka, "Chemometrics approach to drug analysis – An overview," *Am. J. Pharm. Tech. Res.*, vol. 9, pp. 1–13, 2019. doi: https://doi.org/10.46624/ajptr.2019.v9.i1.001
11. "MATLAB and Simulink for Engineered Systems," *mathworks.com*. Available: https://www.mathworks.com/
12. E. Ostertagova, O. Ostertag, J. Kováč, "Methodology and application of the Kruskal–Wallis test," *Appl. Mech. Mater.*, vol. 611, pp. 115–120, 2014. doi: https://doi.org/10.4028/www.scientific.net/AMM.611.115
13. D.F. Specht, "Probabilistic Neural Networks," *Neural Netw.*, vol. 3, pp. 109–118, 1990.
14. B. Mohebali, A. Tahmassebi, A. Meyer-Baese, A.H. Gandomi, "Probabilistic neural networks: a brief overview of theory, implementation, and application," *Handbook of probabilistic models*, pp. 347–367, 2020. doi: https://doi.org/10.1016/B978-0-12-816514-0.00014-X
15. Y. Zeinali, B.A. Story, "Competitive probabilistic neural network," *ICAE*, vol. 24(2), pp. 105–118, 2017. doi: https://doi.org/10.3233/ICA-170540
16. J. Miller, J.C. Miller, *Statistics and chemometrics for analytical chemistry*. Pearson education, 2018.

Received 15.06.2024

INFORMATION ON THE ARTICLE

Yaroslava M. Pushkarova, ORCID: 0000-0001-9856-7846, Bogomolets National Medical University, Ukraine, e-mail: yaroslava.pushkarova@gmail.com

Galina M. Zaitseva, ORCID: 0000-0003-3138-6324, Bogomolets National Medical University, Ukraine, e-mail: galinazaitseva777@gmail.com

ПРОГНОЗУВАННЯ МЕХАНІЗМІВ ТОКСИЧНОЇ ДІЇ ФЕНОЛІВ ЗА ДОПОМОГОЮ ЙМОВІРНІСНОЇ НЕЙРОННОЇ МЕРЕЖІ В ПОЄДНАННІ З ТЕСТОМ КРАСКЕЛА–УОЛЛІСА /

Я.М. Пушкарьова, Г.М. Зайцева

Анотація. Прогнозування токсичності хімічних сполук є одним із найважливіших етапів розроблення лікарських засобів. Використання фенольних сполук є перспективним компонентом у фармацевтичній промисловості з багатьма можливими застосуваннями. Працю присвячено застосуванню ймовірнісної нейронної мережі для класифікації 232 фенолів за механізмами їх токсичної дії. Для встановлення впливу молекулярних дескрипторів на достовірну класифікацію фенольних сполук за механізмами їх токсичної дії використали тест Краскела–Уолліса. Показано, що для коректного навчання ймовірнісної нейронної мережі та ефективного прогнозування механізмів токсичної дії фенолів достатньо використовувати лише 5 молекулярних дескрипторів.

Ключові слова: штучна нейронна мережа, класифікація, дизайн ліків, фенол, токсичність.