

Automatic identification of NBOMe illicit psychoactive substances based on combined molecular descriptors

Adelina Ion¹, Mirela Praisler^{1*}, Catalina Mercedes Burlacu¹ and Nicolae Catalin Stanica²

¹"Dunărea de Jos" University, Department of Science and Environment, Domnească Street, 47, Galati, Romania

²"Gh. M. Murgoci" National College, Mathematics Department, Independentei Boulevard Street, 4, Braila, Romania

Abstract. During the last decade, a growing prevalence of new psychoactive substances (NPS) has been noticed by the law enforcement agencies. Although NPS have no medical use due to their very high toxicity, they are often sold on the black market. NBOMe defines a group of toxic amphetamines that has as parent compound 25I-NBOMe, a synthetic derivative of 2C-I (2,5-dimethoxy-4-iodophenethylamine). In this paper, we are presenting a series of Artificial Neural Networks (ANNs) designed to identify the NBOMe class membership based on a mixture of topological and 3D-MoRSE descriptors. For this purpose, the molecular structures of 160 compounds representing NBOMe compounds, narcotics, sympathomimetic amines, potent analgesics, as well as their main precursors have been first optimized. Then a molecular database was formed by computing a large number of topological and 3D-MoRSE descriptors that characterize these structures. This database was used as input for building an ANN system designed to recognize NBOMes. The relevance of the input variables on its classification performance has been assessed and new systems have been built by using different combinations of selected topological and 3D-MoRSE descriptors. The best performing system has been found by comparing various classification efficiency criteria.

1 Introduction

In recent years, a class of potent synthetic psychoactive substances, referred to as dimethoxyphenyl-N-[(2-methoxyphenyl)methyl] ethanolamine derivatives (NBOMes), has been increasingly detected by law enforcement institutions in seizures of controlled substances. These amphetamines, abused for recreational purposes, were initially synthesized for mapping the brain usage of the 2A serotonin receptor. They are very toxic and hence have no medical use [1].

The most frequently seized NBOMe compound is 25I-NBOMe (2-(4-iodo-2,5-dimethoxyphenyl)-N-[(2-methoxyphenyl)methyl] ethanamine) (see Figure 1). This compound is a derivative of 2C-I (2,5-dimethoxy-4-iodophenethylamine) and it is sold by the

* Corresponding author: Mirela.Praisler@ugal.ro

vendors of designer drugs under a variety of names, such as *N-bomb*, *N-bome*, *Smiles*, *25I*, *Bom-25* or *Cimbi-5*. 25I-NBOMe is very potent, being active in sub-milligram doses. Hence, the risk of overdose is very high. Nevertheless, due to its potency and much lower cost than classical psychedelics, it becomes increasingly popular in rave parties [2]. Clinical toxicology studies indicate that consumption of 25I-NBOMe usually generates confusion, panic and anxiety, visual hallucinations, auditory hallucinations, increased heart rate and blood pressure, thought loops, vasoconstriction, nausea, acute kidney injury or oxygen desaturation [3-7].

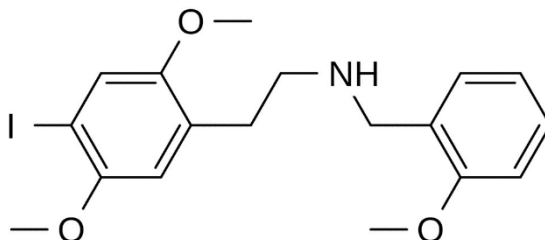


Fig. 1. Skeletal structure of the 25I-NBOMe psychoactive amphetamine.

Artificial neural networks (ANNs) are statistical learning algorithms considered very useful machine learning tools in bioinformatics [8-9]. They simulate the structure and functionality of biological neural networks and have been successfully used for identifying patterns or for classification purposes. ANNs are trained by using algorithms based on the optimization theory and statistical estimation. Real gradients are determined based on the back propagation method and then, by applying a gradient lowering method, the derivative of a cost function is determined in relation to the network parameters. The last procedure determines the best network architecture by modifying the system parameters in a direction related to the gradient [10-12]. This way, ANN are very adequate for analyzing databases that are very large or have incomplete information [13-15].

In this paper, we are presenting and comparing the efficiency of a series of Artificial Neural Networks (ANNs) designed to recognize NBOMe drugs of abuse based on a mixture of topological and 3D-MoRSE descriptors. These molecular descriptors are the result of a mathematical procedure based on the graph theory, which converts a symbolic (2- or 3-dimensional Euclidean) representation of a molecule into a numeric value, i.e. a theoretical descriptor. As these molecular descriptors contain important information about the associated molecular structure, they are very appropriate for describing and classifying substances, as well as for determining correlations between molecular structures and physico-chemical or biological properties.

2 Database and methods

The input database was formed with molecular descriptors calculated for a number of 160 controlled substances that were divided into a class of positives (referred to as *NBOMe*) and one of negatives (*non-NBOMe*). The class of positives contains 15 NBOMe hallucinogens and the class of negatives includes 145 compounds of forensic interest, such as narcotics, sympathomimetic amines, analgesics and their main precursors.

The 3D representation of the molecular structures was obtained with the *HyperChem8.03* software [16] for all the 160 compounds forming the input database. The AM1 semi-

empirical quantum method was applied for the full optimization of their geometries. The geometry was adjusted and the parameters in which the minimum energy of the molecular system is reached were determined based on the Polak-Ribiere mechanism.

A number of 96 topological descriptors and 80 3D-MoRSE descriptors were computed for each of the 160 compounds by using the *Dragon 5.5* software [18]. They have been presented in detailed in a previous published article [19]. The detailed definitions, mathematical formulas and chemical significance of the topological descriptors and 3D-MoRSE (3D-Molecule Representation of Structures based on Electron diffraction) descriptors are detailed in Todeschini et al. [17].

A first ANN, named *176_topo+3D_ANN*, was built with all these 176 molecular descriptors (96 topological descriptors and 80 3D-MoRSE descriptors) by using the Easy NN plus software. The system has three layers (the input, hidden and output layers) and use the sigmoid function as transfer function. The training set consists of 8 NBOME amphetamines, 17 non-NBOME compounds, while the validation set is formed by the remaining 135 substances. The output layer has two output nodes, i.e. NBOME (positives) and non-NBOME (negatives). Convergence was reached for the training process when the average training error falls below the target error of 0.01. The system was trained by using the backpropagation algorithm. Full cross-validation was performed based on the leave-one-out method. The resulting architecture of *176_topo+3D_ANN* consists of 13 hidden nodes and 2314 weight connections.

This initial system was used to evaluate the relative importance of each input descriptors, i.e. its influence on the next layer in the network. Then a new network, named *54_topo+3D_imp_ANN*, was built by including in the database only the first 54 most important descriptors, while using the same sets of compounds and mathematical procedures as for *176_topo+3D_ANN*. This system has 12 hidden nodes and 672 weight connections after optimization.

176_topo+3D_ANN was also used to evaluate the sensitivity of the input descriptors, parameter that indicates how the outputs change when inputs are modified. Then a third system, *43_topo+3D_senz_ANN*, was built with the first 43 most sensitive descriptors. After optimization, this system has 13 hidden nodes and 585 weight connections.

3 Results and discussion

The normalized and relative errors obtained during the training of *176_topo+3D_ANN* is presented in Figure 2. The training process ended after 8 cycles.



Fig. 2. Normalized and relative error obtained during the training of the 176_topo+3D_ANN system, built with 176 molecular descriptors (96 topological descriptors and 80 3D-MoRSE descriptors).

The first 30 most important descriptors, as determined with 176_topo+3D_ANN, are listed in Figure 3, in descending order. We notice that they include mostly 3D-MoRSE descriptors, the most important being Mor18p (signal 18/weighted by polarizability), Mor10u (signal 10/unweighted) and Mor18v (signal 18/weighted by van der Waals volume). In fact, the first 11 most important variables are only 3D-MoRSE descriptors. However, a very positive aspect is that the relative importance is decreasing very slowly.

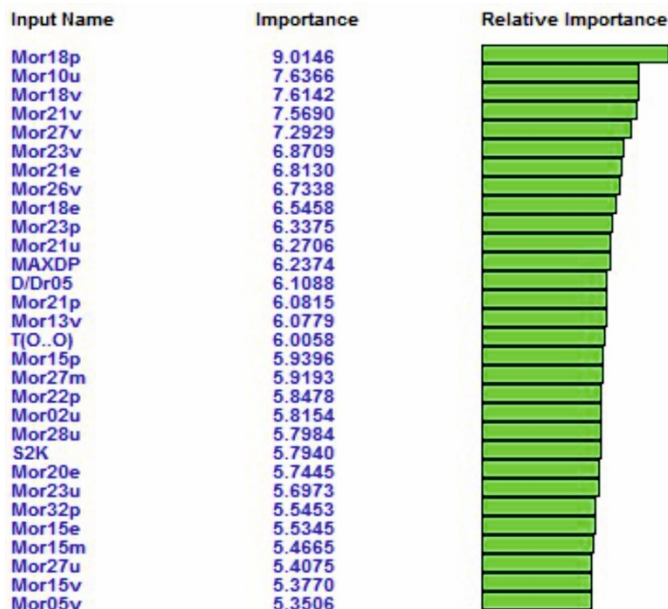


Fig. 3. The first 30 molecular descriptors found to have the highest relative importance by analyzing the 176_topo+3D_ANN system.

The most important topological descriptors are MAXDP (maximal electrotopological positive variation), D/Dr05 (distance/detour ring index of order 5) and T(O..O) (sum of topological distances between O..O).

Figure 4 presents the first 30 descriptors that have the highest relative sensitivity, as assessed by analyzing the 176_topo+3D_ANN system. We notice the same behavior as in the case of the importance analysis. The first 30 descriptors that have the highest relative sensitivity are predominantly 3D-MoRSE descriptors. The most sensitive are Mor20e (signal 20/weighted by Sanderson electronegativity), Mor15p (signal 15/weighted by polarizability) and Mor15v (signal 15/weighted by van der Waals volume). The most sensitive topological descriptors are S2K (2-path Kier alpha-modified shape index, T(O..O)-sum of topological distances between O..O), PHI (Kier flexibility index) and D/Dr06, which is a ring descriptor.

We may notice that the relative sensitivity decreases much faster than the relative importance. Hence, a smaller number of 3D-MoRSE and topological descriptors are worth selecting for building a new ANN system with the most sensitive descriptors.

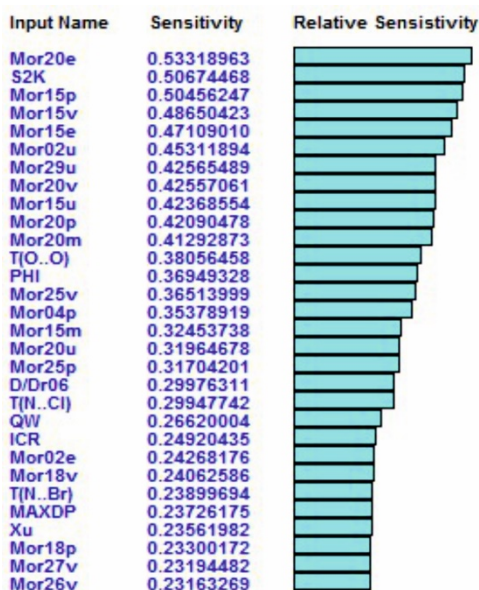


Fig. 4. The first 30 molecular descriptors found to have the highest relative sensitivity by analyzing the 176_topo+3D_ANN system.

In the case of 54_topo+3D_imp_ANN network (which includes 13 topological descriptors and 31 3D-MoRSE descriptors), 18 learning cycles were needed to end the training process, while 43_topo+3D_senz_ANN (built with 13 topological descriptors and 30 3D-MoRSE descriptors) reached convergence after only 5 learning cycles (see Figure 5 and Figure 6).

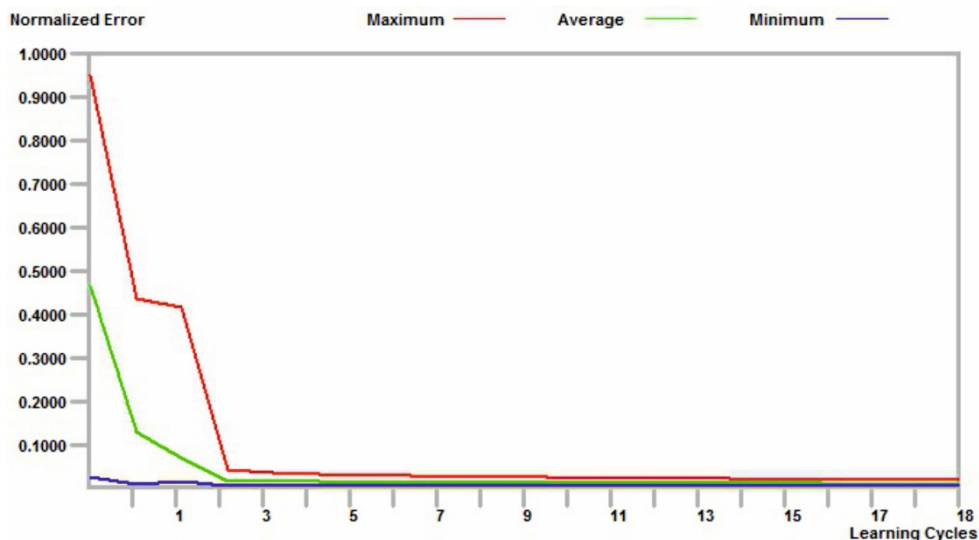


Fig. 5. Training process of the 54_topo+3D_imp_ANN system, built with the first 54 most important descriptors, as determined with the 176_topo+3D_ANN system.

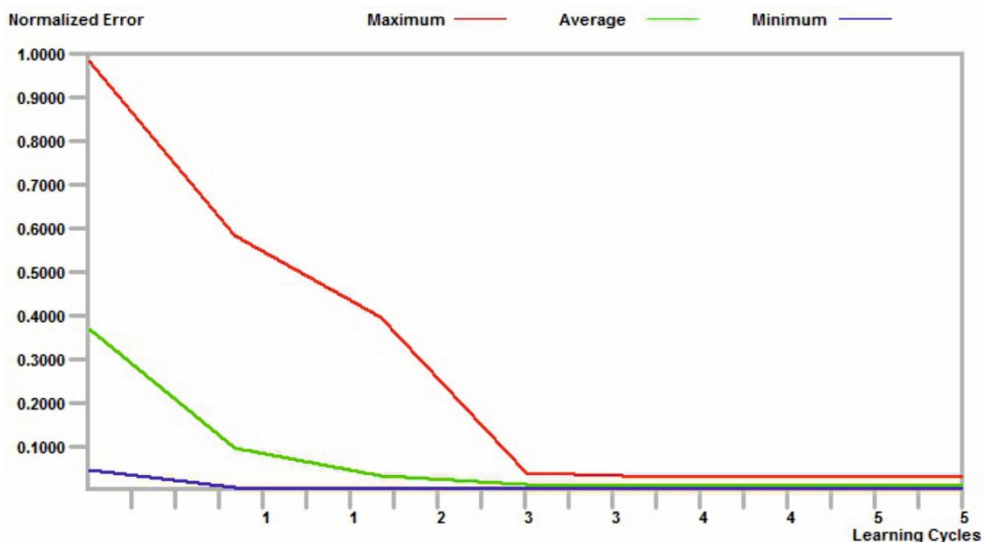


Fig. 6. Training process of the 43_topo+3D_senz_ANN system, built with the first 43 most sensitive descriptors, as determined with the 176_topo+3D_ANN system.

The efficiency of the ANN system in identifying the class identity of an unknown sample has been evaluated based on several figures of merit, i.e. the rate of true positives (TPR), of true negatives (TNR), of false positives (FNR), the classification rate (CR) and the correct classification rate (CCR), as presented in Table 1. They indicate that all ANNs are very efficient, being characterized by remarkably good figures of merit.

Table 1. Results of the validation process.

| | 176_topo+3D_A NN | 54_topo+3D_imp_ ANN | 43_topo+3D_senz_ ANN |
|----------------|-----------------------------|--------------------------------|---------------------------------|
| TPR (%) | 100 | 100 | 100 |
| TNR (%) | 92.36 | 96.55 | 95.86 |
| FNR (%) | 0 | 0 | 0 |
| FPR (%) | 7.64 | 3.45 | 4.18 |
| CR (%) | 99.38 | 100 | 100 |
| CCR (%) | 93.08 | 96.87 | 96.25 |

The results indicate that all the ANNs presented in this study are exceptionally sensitive, as they are all detecting the NBOME psychotropic drugs without exception (TPR = 100%, FNR = 0%). The systems are also very sensitive from the point of view of the negatives, the proportion of actual negatives recognized as such being also very good (TNR \geq 92.36%) for all networks. The best results are obtained with 54_topo+3D_imp_ANN (TNR = 96.55%).

Only very few negatives are misclassified as NBOMes (FPR \leq 7.64%). This figure of merit indicates best that the selection of variables based on their sensitivity, and especially on their importance, is a very useful step for optimizing the efficiency of these screening systems. The FPR obtained for 54_topo+3D_imp_ANN, the system built with the most important descriptors is less than half of the FPR rate obtained for the 176_topo+3D_ANN system.

In addition, the systems built with selected variables have a better capacity to assign a class identity to the analyte than 176_topo+3D_ANN. Both 54_topo+3D_imp_ANN and 43_topo+3D_senz_ANN were able to classify all the samples, while 176_topo+3D_ANN missed on sample (CR = 99.38%). Out of the classified samples, the systems built with selected descriptors classify correctly the same percentage of the samples, which is 3% higher than the CCR of 176_topo+3D_ANN.

4 Conclusions

The detection of NBOME designer drugs is extremely important in forensic practice. The most important characteristic of any system screening for these drugs of abuse is its capacity of recognizing the positive samples, which should not be missed under any circumstance. The results of this study show that the 54_topo+3D_imp_ANN system is the most efficient system screening for NBOME. This artificial intelligence application can be used successfully to predict and estimate the toxicity of any novel compounds having a molecular structure similar to NBOME psychotropic drugs of abuse. This way, it may save the high costs of analytical and toxicological studies.

In addition, we should underline the benefits brought by mixing different types of molecular descriptors, in our case topological descriptors and 3D-MoRSE descriptors. This approach has led to selections of variables having higher importance / sensitivity than if the selections would have been made only with the same number of most important descriptors (54) or most sensitive ones (43) of the same type (only topological or only 3D-MoRSE descriptors). This is certainly one of the main reasons for the significant improvement of the

classification efficiency noticed for the two ANN systems built with selections of variables in comparison with 176_topo+3D_ANN.

References

1. J. Suzuki, M. A. Dekker, E. S. Valenti, et al., *Psychosomatics* **56**, 129-139 (2015)
2. https://www.deadiversion.usdoj.gov/drug_chem_info/nbome.pdf
3. S. L. Hill, T. Doris, S. Gurung, S. Katebe, A. Lomas, M. Dunn, P. Blain, S. H. L. Thomas, *Clin Toxicol* **51**, 487-492 (2013)
4. A. Kelly, B. Eisenga, B. Riley, B. Judge, *Clin Toxicol (Phila)* **50**, 702 (2012)
5. J. Suzuki, J. L. Poklis, A. Poklis, *J. Psychoactive Drugs* **46**, 379-382 (2014)
6. B. V. Dean, S. J. Stellpflug, A.M. Burnett, K. M. J. Engebretsen, *Med Toxicol* **9**, 172-178 (2013)
7. European Monitoring Centre for Drugs and Drug Addiction, Joint EMCDDA–Europol Publications, *Report on a new psychoactive substance: 25I-NBOMe (4-iodo-2,5-dimethoxy-N-(2-methoxybenzyl)phenethylamine)* (Publications Office of the European Union, Luxembourg, 2014)
8. E. Pasomsub, C. Sukasem, S. Sungkanuparph, B. Kijisirikul, W. Chantratita, *Jpn J Infect Dis* **63**, 87-94 (2010)
9. W. Ze, X.-C. Li, W.-X. Zhu, Yao Xue Xue Bao **41**, 1180-1183 (2006)
10. S. Haykin, *Neural Networks and Learning Machines*, 3rd edn. (Pearson Education, New York, 2008)
11. Z. Zhang, *Artificial Neural Network. In: Multivariate Time Series Analysis in Climate and Environmental Research* (Springer, Cham, 2018).
12. M. Karelson, D. Dobchev, *Expert Opin Drug Discov* **6**, 783-796 (2011)
13. S. Gosav, M. Praisler, D.O. Dorohoi, *J Mol Struct* **834**, 188-194 (2007)
14. S. Gosav, M. Praisler, M. L. Birsa, *Int J Mol Sci* **12**, 6668-6684 (2011)
15. S. Gosav, M. Praisler, *Rom Rep Phys* **54**, 929-935 (2009)
16. Hyperchem software, Version 8.0.3., Hyper Co., USA, 2007.
17. R. Todeschini, V. Consonni, *Handbook of molecular descriptors* (Wiley, Weinheim, 2000)
18. Talete srl, DRAGON for Windows (Software for Molecular Descriptors Calculation), Version 5.5 -2007 , Milano – Italy, <http://www.talete.mi.it/>
19. A. Ion, S. Gosav, M. Praisler, C. Negoita, *Artificial neural network designed to identify NBOMe hallucinogens based on 3D-MoRSE descriptors and topological descriptors*, in Precup R.-E. (Ed.), *ICSTCC 23rd International Proceedings* **8885908**, 872-876 (2019)