ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
MATHEMATICS, PHYSICS, THEORETICAL MECHANICS
FASCICLE II, YEAR X (XLI) 2018, No. 1

Article DOI: https://doi.org/10.35219/ann-ugal-math-phys-mec.2018.1.05

ARTIFICIAL NEURAL NETWORK DESIGNED TO IDENTIFY NBOMe HALLUCINOGENS BASED ON MOLECULAR DESCRIPTORS

Steluţa Gosav[1], Adelina Ion[1], Mirela Praisler[*1]

[1]Faculty of Science and Environment,"Dunărea de Jos" University of Galati, Domnească Street, 47, RO-800008, Galati, Romania
[*] Corresponding author: Mirela.Praisler@ugal.ro

Abstract
In the last few years there has been a rapid increase in the availability and recreational use of synthetic hallucinogens. One novel group of toxic phenethylamine derivatives, referred to as NBOMe, has recently gained prominence. The goal of this study was to develop an Artificial Neural Network (ANN) able to classify NBOMe hallucinogens based on their molecular descriptors. The database consists of 161 compounds representing drugs of abuse (NBOMe hallucinogens, sympathomimetic amines, narcotics and other potent analgesics), precursors, or derivatized counterparts. The molecular structures of all the compounds included in the database have been first optimized and then the molecular descriptors have been determined by using the *Dragon 5.5.* software. The validation has been performed by using all the available samples and the leave-one-out algorithm. The efficiency with which the ANN system identifies the class identity of an unknown sample was evaluated by calculating several figures of merit.

Keywords: NBOMe hallucinogens, molecular descriptors, ANN.

1. INTRODUCTION

The molecular structures of phenethylamines contain a phenyl ring, joined to an amino group via an ethyl side chain (see Figure 1). These designer drugs, which are also called „N-bomb", „legal acid", „smiles" or "25I", are highly potent hallucinogens that are regarded as alternatives to LSD [1].

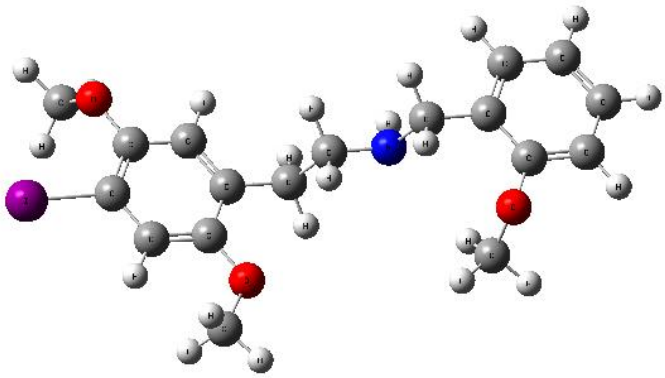

Figure 1. Molecular structure of the 25I-NBOMe hallucinogenic phenethylamine.

Artificial Neural Networks (ANNs) are a group of statistical learning algorithms. They are derived from a simplified brain concept and consist of a computer-based system in which a number of nodes, called processing elements or neurons, are interconnected in a network-like structure. Thanks to their adaptive nature, they are powerful pattern recognition methods. They may also be considered machine learning techniques [2-3].

Many algorithms are available for training the neural network models. They are usually based on the optimization theory and statistical estimations. Backpropagation is often used to compute the actual gradients. Then, some form of gradient descent methodology is applied by simply taking the derivative of the cost function with respect to the network parameters. Then the system changes these parameters in a gradient-related direction [4-6].

The most frequently used architecture of an ANN is the multilayer feed-forward network, i.e. a network in which the nodes are divided into three types of layers. The first layer consists of the input nodes and it is followed by one or more layers of hidden nodes. The last layer consists of output nodes. The input nodes activities and the weights are associated with the connections between the nodes. The hidden nodes determine the activity of each hidden node.

ANNs have the ability to learn from experience, to improve performance and adapt to changes in environment. Also, they are able to handle incomplete information. They may be very effective, especially in situations where it is not possible to identify rules or steps necessary for solving a problem [7-9].

2. EXPERIMENTAL PART

The database consists of 160 compounds representing forensic substances such as drugs of abuse (NBOMe hallucinogens, sympathomimetic amines, narcotics and other potent analgesics), precursors and derivatized counterparts, out of which 15 are NBOMe hallucinogens (class code NBOMe) and 145 other types of hallucinogens (class code non-NBOMe).

The molecules included in the database have been represented in 3D coordinates by using the *HyperChem 8.03* software package [10]. Their molecular geometries have been optimized by using the AM1 semi-empirical quantum mechanics method, the Polak-Ribiere algorithm being applied to adjust the geometry and determine the conditions in which the minimum energy of molecular system is reached.

A number of 74 constitutional descriptors and functional groups (CD+FG) were computed for each of the 160 compounds included in the database, by using the *Dragon 5.5* software package. The calculated descriptors are summarized in Table 1.

Table 1. List of constitutional descriptors and functional group counts.

Description	Symbols	Description	Symbols
Average molecular weight	AMW	Number of ring secondary C(sp3)	nCrs
Molecular weight	MW	Number of ring tertiary C(sp3)	nCrt
Sum of atomic van der Waals volumes	S_v	Number of ring quaternary C(sp3)	nCrq
Sum of atomic Sanderson electro negativities	S_e	Number of aromatic C(sp2)	nCar
Sum of atomic polarizabilities	S_p	Number of unsubstituted benzene C(sp2)	nCbH
Sum of Kier-Hall electro topological states	S_s	Number of substituted benzene C(sp2)	nCb-
Mean atomic van der Waals volume	M_v	Number of non-aromatic conjugated C(sp2)	nCconj
Mean atomic Sanderson electro negativity	M_e	Number of aliphatic secondary C(sp2)	nR=Cs

Mean atomic polarizability	M_p	Number of esters (aliphatic)	nRCOOR
Mean electro topological state	M_s	Number of esters (aromatic)	nArCOOR
Number of molecule atoms	nAT	Number of secondary amides (aliphatic)	nRCONHR
Number of non-H atoms	nSK	Number of tertiary amides (aliphatic)	nRCONR2
Number of bonds	nBT	Number of primary amines (aliphatic)	nRNH2
Number of non-H bonds	nBO	Number of secondary amines (aliphatic)	nRNHR
Number of multiple bonds	nBM	Number of tertiary amines (aliphatic)	nRNR2
Number of double bonds	nDB	Number of imides (-thio)	nN(CO)2
Number of aromatic bonds	nAB	Number of hydroxyl groups	nROH
Number of heavy atoms	nHM	Number of secondary alcohols	nOHs
Atom-type counts	nH, nC, nN, nO, nS, nF, nCL, nBR, nI, nX	Number of ethers (aromatic)	nArOR
Functional group counts	nOH, nNH, nNH2, nCO	Number of sulfides	nRSR
Number of donor atoms for H-bonds	nHD	Number of X on aromatic ring	nArX
Number of acceptor atoms for H-bonds	nHA	Number of donor atoms for H-bonds (with N and O)	nHDon
Counts of different size rings	nCIC, nCIR, nR04, nR05, nR06, nR07, nR08, nR09, nR10, nR11, nR12	Number of acceptor atoms for H-bonds (N O F)	nHAcc
Sum of conventional bond orders (H-depleted)	SCBO	Number of Pyrrolidines	nPyrrolidines
Aromatic ratio	ARR	Number of Pyridines	nPyridines
Number of rotatable bonds	RBN		
Rotatable bond fraction	RBF		
Number of benzene-like rings	nBnz		
Number of total primary, secondary, tertiary , quaternary C (sp3)	nCp, nCs, nCt, nCq		

The ANN was built using the *Easy NN plus* software and has three layers (input, hidden and output layer). The transfer function is the sigmoid function. The network was trained by using the backpropagation algorithm. The training set consists of 8 NBOMe hallucinogens and 17 non-NBOMe compounds. The validation set is formed by the remaining 135 samples. The neural network (74CD-GF_ANN) has as input variables a number of 74 constitutional descriptors and functional groups that have been calculated for each sample. Its architecture, which resulted from the optimization process, consists of 12 hidden nodes and 912 weight connections. The 74CD-GF_ANN system was designed with two output nodes, i.e. NBOMe and non-NBOMe compounds (see Fig.2).

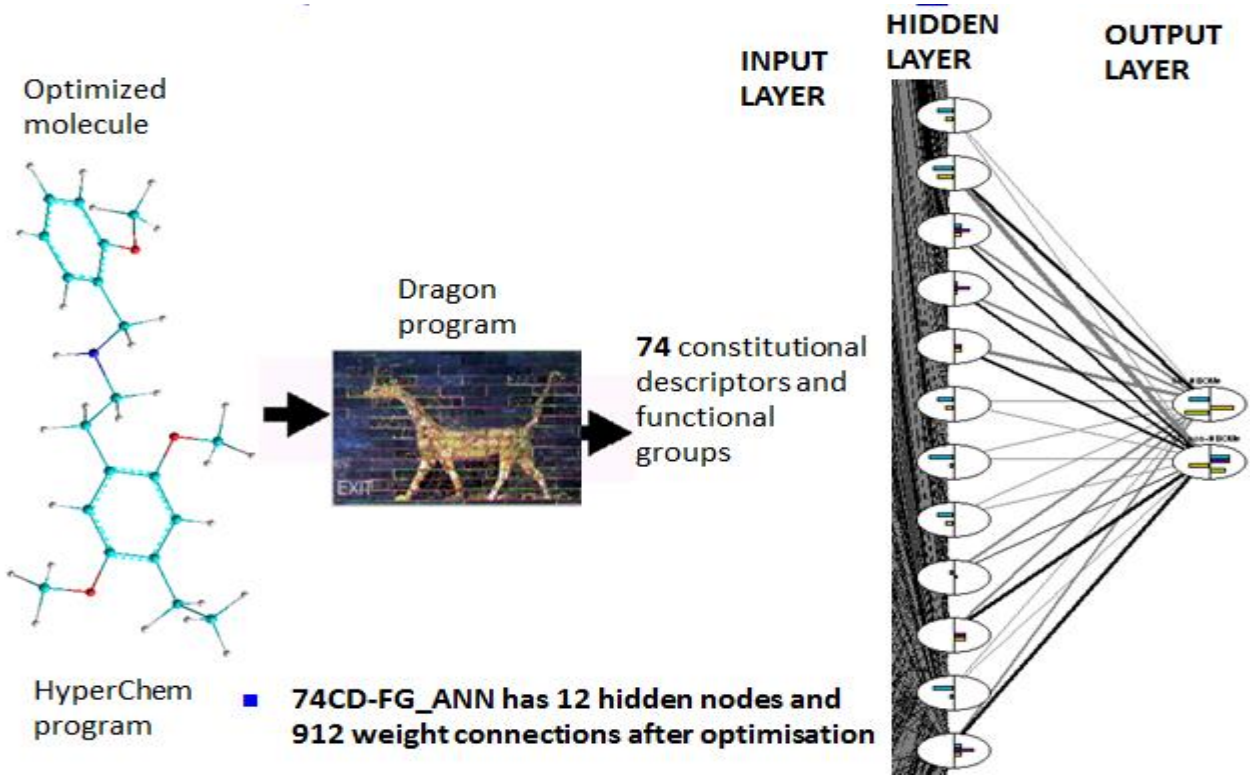

Figure 2. Development of the 74CD-FG_ANN system.

3. RESULTS AND DISCUSSION

The neural networks has been programmed to stop the training process when the training average error drops below the target error set at TE = 0.01. The process is illustrated in Fig. 3.

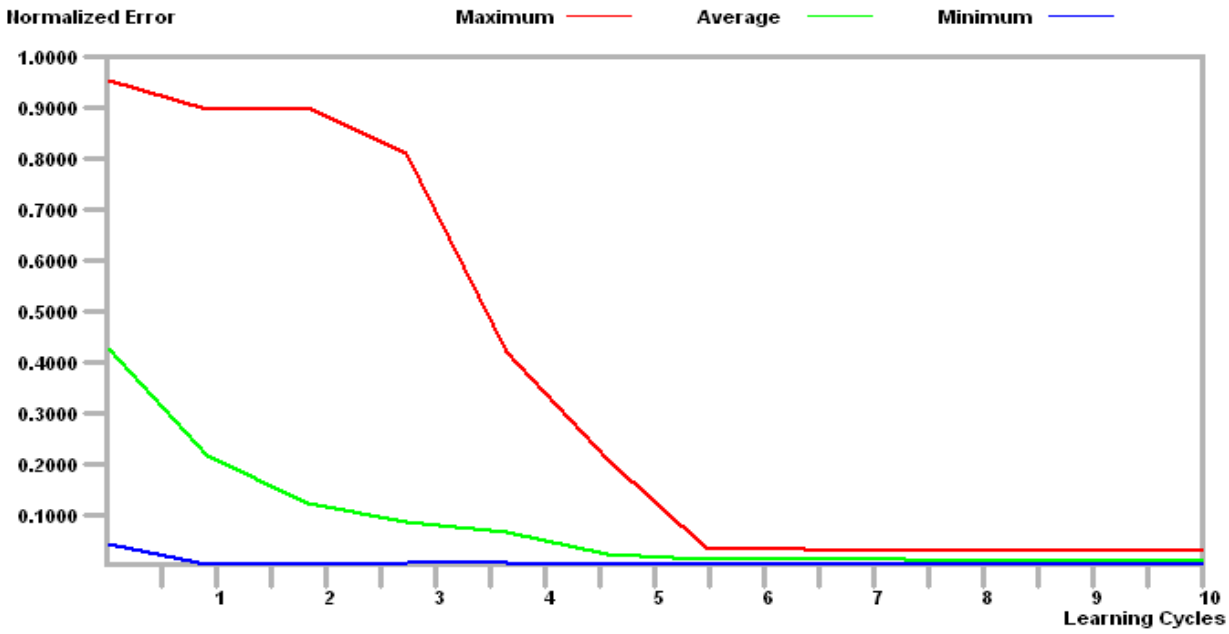

Figure 3. Evolution of the training process of the 74CD-GF_ANN neural network.

The validation was performed by applying the leave-one-out algorithm for all the remaining samples. The efficiency with which 74CD-GF_ANN identifies the class identity of an unknown sample was evaluated by calculating several figures of merit: the rate of true positives (TP), of true negatives (TN), of false positives (FP), of false negatives (FN), of classification (C), and of correctly classified samples (CC).

Table 2. The results of the validation process.

	74CD-FG_ANN
TP (%)	100.00
TN (%)	80.99
FN (%)	0.00
FP (%)	19.01
C (%)	98.13
CC (%)	82.80

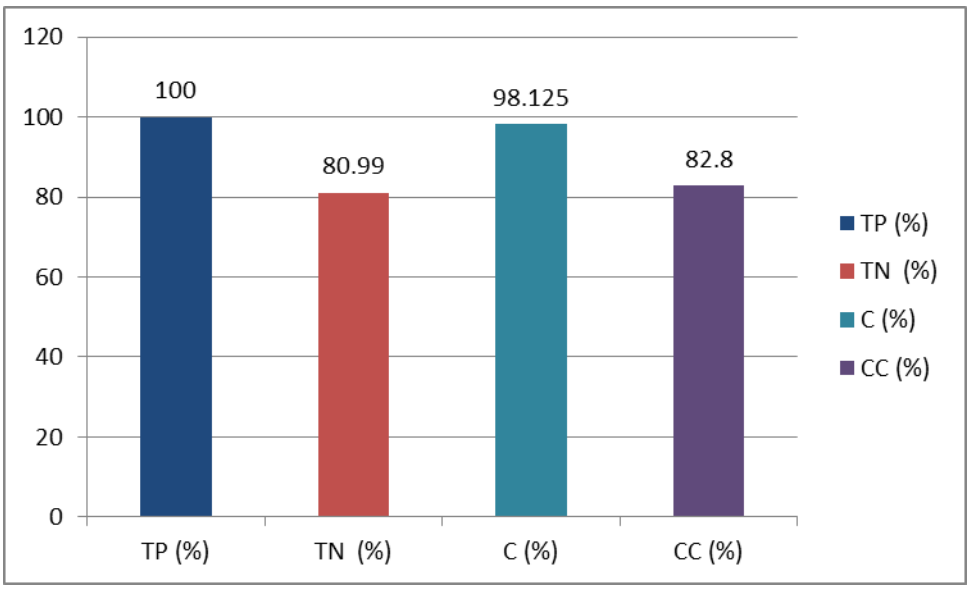

Figure 4. The rate of true positives (TP), of true negatives (TN), of classification (C) and of correctly classified samples (CC) for the 74CD-FG_ANN system.

The absolute importance of an input variable (node) was determined as the sum of the absolute weights characterizing the connections between this input node and the nodes of the hidden layer. The input importance is a measure of how each input will influence the next layer in the network. Fig. 5 illustrates the relative importance of the most efficient of the 74 constitutional descriptors and functional group counts.

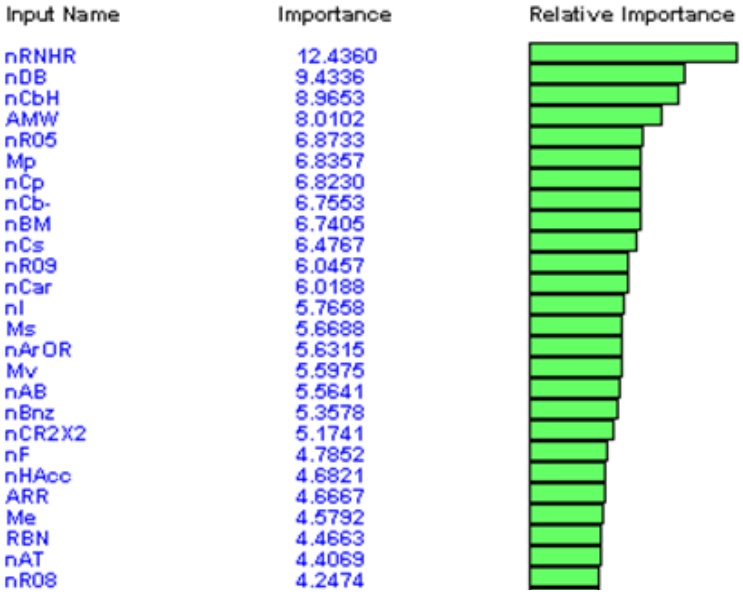

Figure 5. Variables found to have the highest relative importance for the 74CD-FG_ANN system.

The relative sensitivity is a measure of how much the outputs change when the inputs are changed. The inputs are all set to the lowest values and then each in turn is increased to the highest value. The change in the outputs is measured as each input is increased from lowest to highest in order to establish the sensitivity to change. As Fig. 6 shows, the 74CD-FG_ANN system has a very good sensitivity, which means that it recognizes the NBOMe hallucinogens very efficiently.

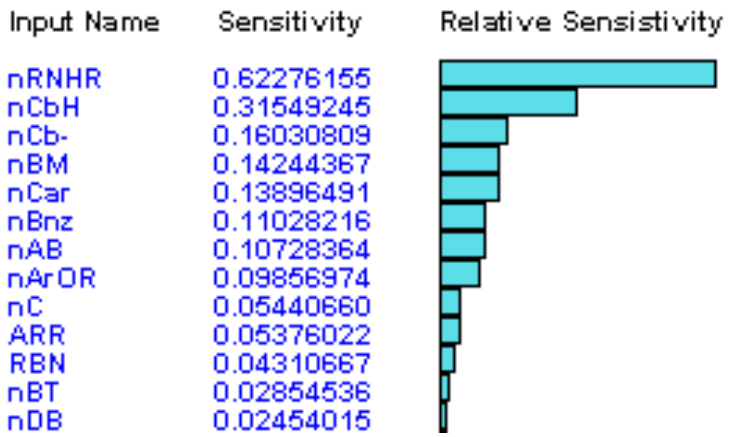

Figure 6. Relative sensitivity of the 74CD-FG_ANN system.

4. CONCLUSIONS

The 74CD-FG_ANN system, built with constitutional descriptors and functional groups, yields very good results. It allows the identification of any of the hallucinogens included in the database, as well as of any unknown (new drug of abuse not included in the database) that has a biological activity and toxicity similar to the NBOMe class of illicit hallucinogenic amphetamines. An important advantage of this detection process is that this type of detection avoids the costs related to the synthesis, as well as of the clinical and toxicological tests that must be performed to evaluate the toxicity of the unknowns found on the black market.

If 74CD-FG_ANN classifies a compound as positive, it means that the compound has the same biological activity as the modeled NBOMe hallucinogens. The correct classification of the positives is especially important in the forensic practice, where the positive identification of drugs of abuse must not fail.

In our case, the 74CD-FG_ANN system has a very good sensitivity (a TP rate of 100%), meaning that all positives (NBOMe hallucinogens) are correctly recognized as such. This also means that the descriptors that were chosen to build this network contain enough of the discriminant information needed to distinguish NBOMe hallucinogens from negatives. The selectivity, as measured by the efficiency in classifying the negatives (non-NBOMe hallucinogens) is also very good (TN = 80.99%).

References

1. M.F. Weaver, J.A. Hopper and E.W. Gunderson, Designer drugs 2015: assessment and management, Weaver et al. *Addiction Science & Clinical Practice* (2015).
2. E. Pasomsub, C. Sukasem, S. Sungkanuparph, B. Kijsirikul, W. Chantratita, The Application of Artificial Neural Networks for Phenotypic Drug Resistance Prediction: Evaluation and Comparison with Other Interpretation Systems. *Jpn. J. Infect. Dis.*, 2010, 63(2), 87-94.
3. W. Ze, X.-C. Li, W.-X Zhu, Prediction of drug bioavailability by genetic algorithm and artificial neural network. *Yaoxue Xuebao*, 2006, 41(12), 1180-1183.
4. Y.C. Sun,. Y.X. Peng, Y.X. Chen, A.J. Shukla, Application of artificial neural networks in the design of controlled release drug delivery systems. *Adv. Drug Deliv. Rev.*, 2003, *55*(9), 1201-1215.
5. E. Byvatov, U. Fechner, J. Sadowski, G. Schneider, Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *J. Chem. Inf. Comput. Sci.*, 2003, *43*(6), 1882-1889.
6. M. Karelson, D. Dobchev, Using artificial neural networks to predict cell-penetrating compounds. *Expert Opin. Drug Discov.*, 2011, *6*(8), 783-796.
7. S. Gosav, M. Praisler, D.O. Dorohoi, G. Popa, Structure – Activity Correlations for Illicit Amphetamines Using ANN and Constitutional Descriptors, *Talanta*, 2006, 70, 922-928.
8. S. Gosav, M. Praisler, M. L. Birsa, Principal Component Analysis Coupled with Artificial Neural Networks—A Combined Technique Classifying Small Molecular Structures Using a Concatenated Spectral Database, *International Journal of Molecular Sciences*, 2011, 12-10, 6668-6684.
9. S. Gosav, M. Praisler, M.L. Birsa, Principal Component Analysis Coupled with Artificial Neural Networks—A Combined Technique Classifying Small Molecular Structures Using a Concatenated Spectral Database, *International Journal of Molecular Sciences*, 2011, 12, 6668-6684.
10. *Hyperchem software*, Version 8.0.3., Hyper Co., USA, 2007.