

ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
MATHEMATICS, PHYSICS, THEORETICAL MECHANICS
FASCICLE II, YEAR XI (XLII) 2019, No. 2
DOI: <https://doi.org/10.35219/ann-ugal-math-phys-mec.2019.2.06>

Choosing relevant functional groups for optimizing Artificial Neural Networks detecting NBOMe hallucinogens

Adelina Ion¹, Steluța Gosav^{1,2}, Mirela Praisler¹

¹*Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment,
„Dunărea de Jos” University of Galati, Romania*

²*INPOLDE interdisciplinary research network, ReForm multidisciplinary Platform, Dunarea de Jos
University of Galati, Faculty of Sciences and Environment, 111 Domneasca St., 800201 Galati, Romania,
e-mail: Adelina.Ion@ugal.ro, Steluta.Gosav@ugal.ro, Mirela.Praisler@ugal.ro*

Abstract

In the early 2010s, a new group of illicit psychedelic phenethylamines was reported by the law enforcement agencies, namely the NBOMe hallucinogens. The latter seem to be sold on the black market as an alternative to LSD, due to their powerful psychoactive effects. The goal of this study was to develop an optimized Artificial Neural Network (ANN) able to classify NBOMe hallucinogens based on their functional groups. These chosen molecular descriptors (functional groups) have been computed, by using the *Dragon 5.5* program, for the molecular structures of the main NBOMe hallucinogens, which have been first optimized by using the *Hyperchem* program. The ANN system was built with the *Easy NN plus* program. Then, the importance of each functional group has been assessed. A new input database has been built with the functional groups found to be the most important. The performance of the new ANN system has been characterized based on several classification accuracy criteria. The impact of the variable selection on the ANN performances is discussed in detail.

Keywords: NBOMe psychedelic phenethylamines, functional groups, Artificial Neural Networks.

1. INTRODUCTION

New and harmful psychoactive substances are emerging at an unprecedented rate, so illicit drugs pose a serious threat to the health and safety of people around the world. A novel class of potent synthetic hallucinogens (NBOMe hallucinogens), originally developed for research purposes, was reported by the law enforcement agencies.

25I-NBOMe is a derivative of 2C-I (4-iodo-2,5-dimethoxyphenethylamine), which is the most representative compound of NBOMe hallucinogens. It is sold on the black market as an alternative to LSD, due to its powerful psychoactive effects [1]. Its chemical structures is represented in Fig. 1.

An Artificial Neural Network (ANN) is an artificial intelligence application designed to simulate the mode in which the human brain processes information. This computer-based system learns through experience, not from programming. It gathers the necessary knowledge by detecting patterns in the data. It is an excellent tool for classification purposes [2-3].

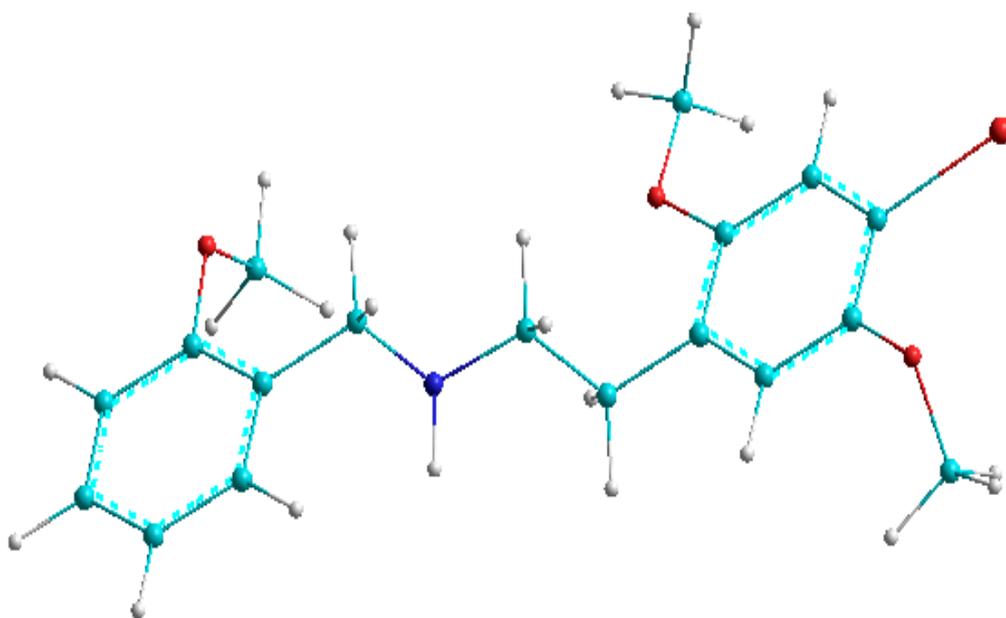


Fig. 1. Molecular structure of the 25I-NBOMe hallucinogenic phenethylamine.

The most widely used ANN is a supervised network, that consist of three or more layers. One or more hidden layers, which consist of nodes capturing nonlinearity in the data, are situated between the input layer (consisting of the input nodes) and the output layer (formed by the output nodes)

Most of the time, the backpropagation learning rule is used for calculating real gradients. This method refers to how the error generated at the output layer is propagated back to the hidden layer and finally to the input layer. The activation of a hidden layer node is calculated by using a transfer function. Usually, the sigmoid function is used in this procedure [4-6].

The network reaches the highest level of accuracy during the training process when the connections between units are optimized until the prediction error is reduced below a set limit [7].

The goal of this study was to develop a series of ANNs able to classify NBOMe hallucinogens based on their most important functional groups. The size of the input database was varied in order to find the most effective architecture.

2. EXPERIMENTAL

The input database was created by calculating the molecular descriptors characterizing 160 controlled substances, representing illicit drugs (hallucinogens NBOMe, narcotics, stimulant amphetamines and potent analgesics), precursors and derived homologs. The database includes 15 NBOMe hallucinogens, which have been assigned the class code "NBOMe". The remaining 145 substances have been assigned the "non-NBOMe" class code.

The molecular structures of all the 160 compounds were represented in 3D coordinates by using the *HyperChem8.03* software package [8]. The geometry of the molecular structures was optimized based on the semi-empirical AM1 method. The geometry was adjusted by applying the Polak-Ribere mechanism, until the conditions of the minimum energy of the molecular system have been reached.

These chosen molecular descriptors (functional groups) have been computed, for the molecular structures of the main NBOMe hallucinogens, by using the Dragon 5.5 program. An ANN was created, with all these 31 functional groups, by using the *Easy NN plus* software. The most important descriptors were identified from this network and then, two new ANNs have been constructed by including various numbers of important descriptors in the database.

All these networks have the same architecture, which consists of three layers (input, hidden and output layers). The transfer function used was the sigmoid function. The backpropagation algorithm has been applied in order to train the systems. These ANNs have been constructed by using descriptors calculated for the same sets of compounds. More specifically, the training set contains 8 NBOMe and 17 non-NBOMe hallucinogens. The validation set consists of the remaining 135 samples.

The ANN training process is completed when convergence is reached, more precisely when the average training error falls below the target error. The threshold was set at $TE = 0.01$ for all networks. The systems have been cross-validated based on the leave-one-out algorithm. Then, the efficiency of each ANN network has been assessed based on its ability to correctly identify the class identity of an unknown sample.

The 31FG-ANN system has as input variables the 31 functional groups that have been computed for each of the compounds mentioned above. In order to estimate the minimum number of input variables needed to obtain an efficient system, another two ANNs have been built by using as input variables the first 20 most important functional groups (20imp_FG_ANN network) and the first 10 most important functional groups (10imp_FG_ANN network) of 31FG-ANN, respectively.

The relative importance, determined for the first 20 most important descriptors of the 31FG-ANN network in descending order, is shown in Fig. 2.

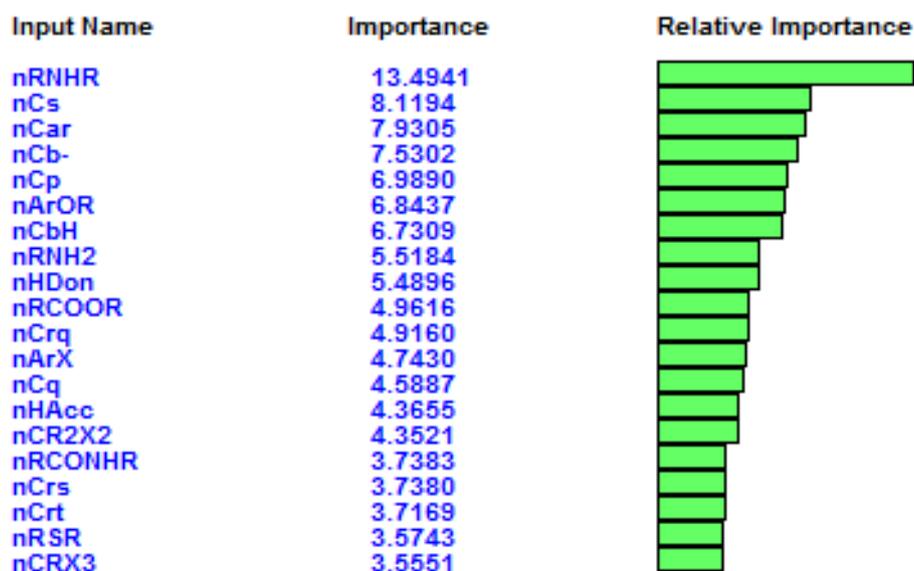


Fig. 2. The first 20 variables found to have the highest relative importance by analyzing the 31FG-ANN artificial neural network.

The 20imp_FG_ANN network has 11 hidden nodes and 242 weight connections after optimization. A number of 15 learning cycles were needed for ending the learning process (see Fig. 3).

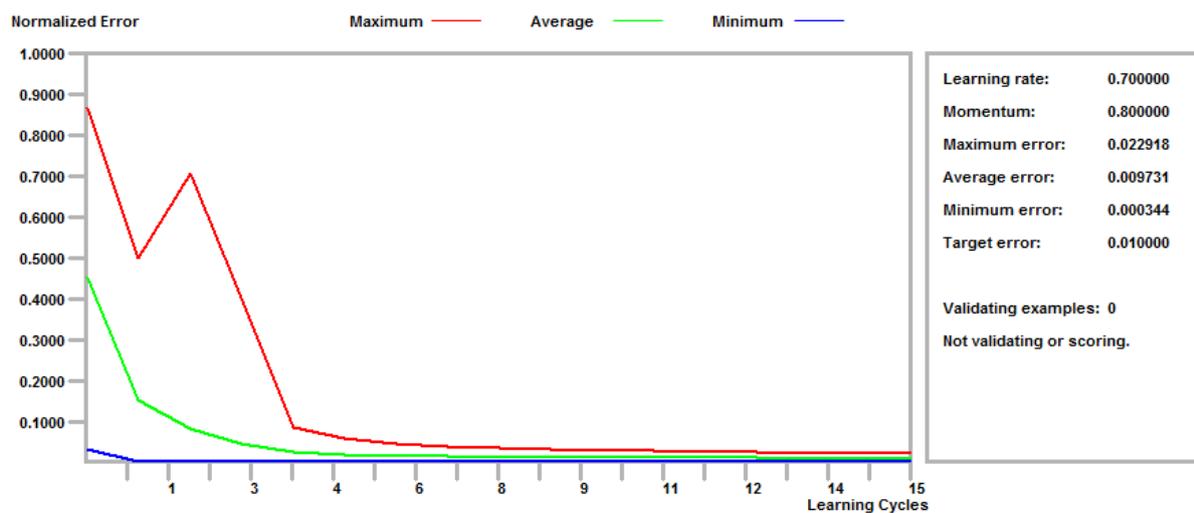


Fig. 3. Training the 20imp_FG_ANN artificial neural network, built with the first 20 most important descriptors.

Only 12 learning cycles were needed to train the ANN built with only 10 most important functional groups (10imp_FG_ANN network). This system has 8 hidden nodes and 96 weight connections after optimization. The architecture of 10imp_FG_ANN is illustrated in Fig. 4.

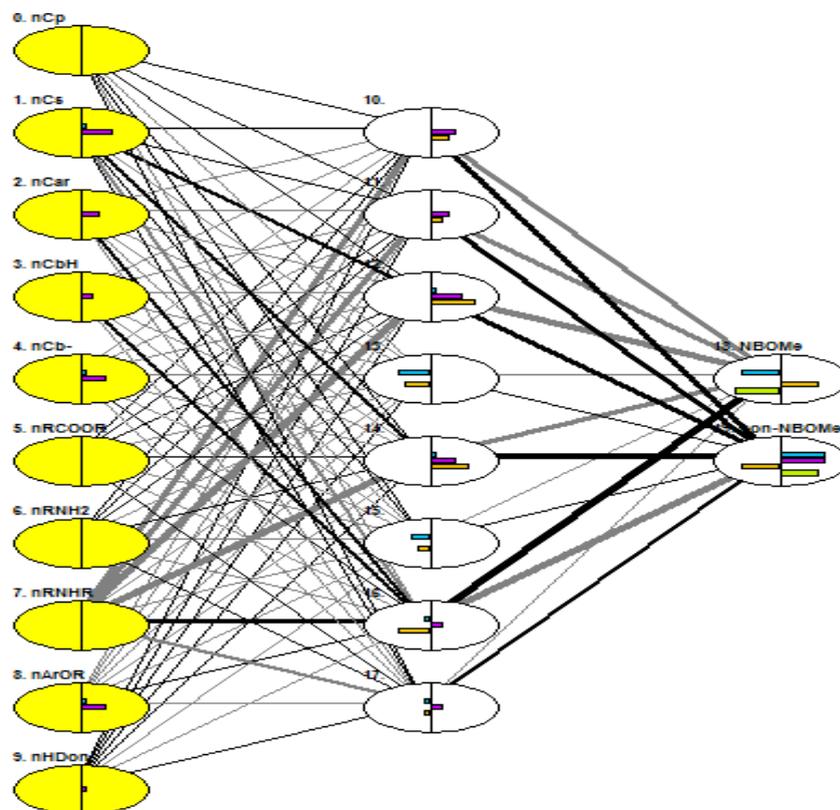


Fig. 4. The architecture of the 10imp_FG_ANN artificial neural network, built with the first 10 most important functional groups identified with the 31FG-ANN artificial neural network.

3. RESULTS AND DISCUSSION

In order to compare the classification efficiency of these ANNs, several figures of merit have been calculated after the validation process of each ANN: the rate of true positives (TPR), of true negatives (TNR), of false positives (FPR), of false negatives (FNR), of classification (CR), and accuracy (ACC). The results obtained for the ANN built with the first 20 most important functional groups are presented in Table 1 and those obtained for 10imp_FG_ANN in Table 2.

Table 1. Classification accuracy indicated by the validation process for the 20imp_FG_ANN artificial neural network, which has been built with the first 20 most important functional groups.

Validation parameter	Network
	20imp_FG_ANN
TPR (%)	100
TNR (%)	88.19
FNR (%)	0
FPR (%)	11.80
CR (%)	99.38
ACC (%)	89.31

Table 2. Classification accuracy indicated by the validation process for the 10imp_FG_ANN artificial neural networks built with the first 10 most important functional groups.

Validation parameter	Network
	10imp_FG_ANN
TPR (%)	100
TNR (%)	88.28
FNR (%)	0
FPR (%)	11.72
CR (%)	100
ACC (%)	89.38

The results indicate that both networks fail to recognize as negatives the same false positives, which are N-methyltryptamine, oxyfedrine, P-hydroxiephedrine, pipradol and papaverine. Their chemical structures are represented in Fig. 5.

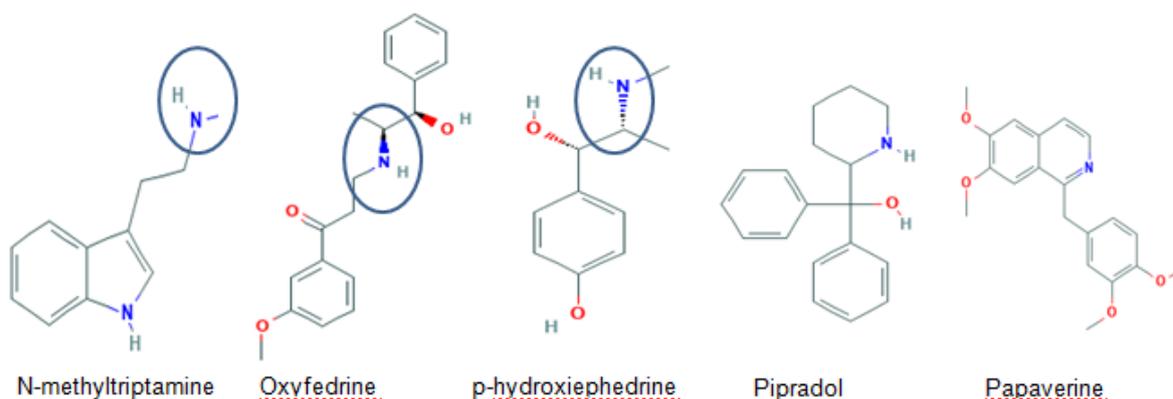


Fig. 5. Molecular structure of the negatives classified as (false) positives by the 20imp_FG_ANN and 10imp_FG_ANN networks.

As Fig. 2 indicates, the most important descriptor is nRNHR, the number of secondary amino groups. We may conclude that the first three compounds have been misclassified as false positives most probably because they contain an aliphatic secondary amine, chemical group that is also found in the molecular structures of the NBOMe hallucinogens. The misclassification of the last two negatives mentioned in Fig. 5, pipradol and papaverine, is probably due to the fact that nCar, the number of aromatic carbons, which is another highly important descriptor (see Fig 2), has the value of 12 (aromatic carbons) for these two negatives as well as for the NBOMe compounds.

4. CONCLUSIONS

Table I and Table II indicate that both ANN systems (20imp_FG_ANN network and 10imp_FG_ANN) created with only the most important functional groups, as identified with the 31FG_ANN network, have a remarkable sensitivity. They are both characterized by TPR = 100%, which means that all positives (NBOMe hallucinogens) are recognized as such. Hence, if the input variables are selected based on their importance, the input database can be diminished from 31 to 10 functional groups without jeopardizing the sensitivity of the system screening for NBOMe hallucinogens.

Both 20imp_FG_ANN network and 10imp_FG_ANN networks have also a good selectivity, their correct classification rate of the negatives (TNR > 86%) being nearly the same. Hence, selecting the variables based on their importance does not affect their capacity to distinguish NBOMe hallucinogens from negatives either.

An important observation is that the networks created with the most important functional groups were able to classify all the samples subjected to classification, with the exception of 20imp_FG_ANN, for which CR = 99.38%. On the other hand, decreasing the number of input variables (according to their importance) to 10, increased CR up to 100% (for 10imp_FG_ANN).

We may hence conclude that the best performing ANN is 10imp_FG_ANN. In other words, ANNs constructed with selections formed by the most important functional groups, in particular 10imp_FG_ANN, are valuable tools for estimating the toxicity of compounds that have a molecular structure similar to those of the NBOMe hallucinogens. The 10imp_FG_ANN system is also very cost-effective, as it eliminates important costs related to the synthesis, as well as to the clinical and toxicological testing of a unknown drugs that need to be evaluated toxicologically.

References

1. M. F. Weaver, J. A. Hopper, and E. W. Gunderson, Designer drugs 2015: assessment and management, *Addict. Sci. Clin. Pract.* 10, 2015.
2. Y.C. Sun, Y.X. Peng, Y.X. Chen, and A.J. Shukla, Application of artificial neural networks in the design of controlled release drug delivery systems, *Adv. Drug Deliv. Rev.* 55, 1201-1215, 2003.
3. E. Byvatov, U. Fechner, J. Sadowski, and G. Schneider, Comparison of support vector machine and artificial neural network systems for drug/nondrug classification, *J. Chem. Inf. Comput. Sci.* 43, 1882-1889, 2003.
4. S. Gosav, M. Praisler, D.O. Dorohoi, G. Popa, Structure – activity correlations for illicit amphetamines using ANN and constitutional descriptors, *Talanta* 70, 922-928, 2006.
5. S. Gosav, M. Praisler, Artificial neural networks built for the recognition of illicit amphetamines using a concatenated database, *Rom. Rep. Phys.* 54, 929-935, 2009.
6. 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, *Int. J. Mol. Sci.* 12-10, 6668-6684, 2011.
7. S. Gosav, M. Praisler, D.O. Dorohoi, ANN expert system screening for illicit amphetamines using molecular descriptors, *J. Mol. Struct.* 834, 188-194, 2007.