ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
MATHEMATICS, PHYSICS, THEORETICAL MECHANICS
**FASCICLE II, YEAR XII (XLIII) 2020, No. 2**
DOI: https://doi.org/10.35219/ann-ugal-math-phys-mec.2020.2.09

# Improved detection of 2C-x and Dox amphetamines – an analytical tool mitigating the environmental impact of their illicit manufacturing, consumption, and disposal

Cătălin Negoiţă [1], Mirela Praisler[1,2]

[1] *"Dunarea de Jos" University of Galati, Faculty of Sciences and Environment, 47 Domneasca Street, 800008, Galati, Romania,*
*C.Negoita@ugal.ro, Mirela.Praisler@ugal.ro*
[2] *INPOLDE interdisciplinary research network, Dunarea de Jos University of Galati, 111 Domneasca St., 800201 Galati, Romania*

**Abstract**
Numerous studies made worldwide have indicated that treated and untreated wastewater entering surface waters contains illicit drugs, as a result of human consumption and excretion, illicit manufacturing processes, and improper disposal. The results indicate that these wastes, and hence the drugs of abuse exposure, influences the plants and animals and the associated health of streams. The hallucinogenic 2C-x and DOx amphetamines are synthetic psychotropic drugs that have been increasingly reported in seizures, especially in Europe and the US. This paper presents a new combination of artificial intelligence techniques that improves the celerity of the detection of these illicit drugs based on their ATR-FTIR spectra while decreasing the required computational resources.

**Keywords:** generalized additive model, hallucinogens, ATR-FTIR.

## 1. INTRODUCTION

Amphetamines are synthetic substances that have a stimulating role in the central nervous system. They are usually sold in the form of a white powder. The parent compound, amphetamine, is N,α-methylbenzeneethanamine. Because amphetamines are controlled substances, new compounds, which are not on the official list of banned substances, are permanently synthesized in clandestine laboratories [1]. They are obtained by slightly changing the molecular structure that generally has a biological effect similar to the classic amphetamine, but does not fall under the law on prohibited substances. This paper presents a system able to detect automatically two new classes of hallucinogenic analogues of amphetamine, namely 2C-x and DOx. The 2C-x amphetamines have methoxy groups on the 2 and 5 positions of the benzene ring. In the case of the DOx group of amphetamines, the benzene ring is also substituted by an alkyl or a halogen at the 4- position.

Generalized additive models (GAMs) provide a general framework for generalized additive models extending a standard linear model by allowing non-linear functions of each of the variables while maintaining additivity. Just like linear models, GAMs additivity can be applied with both quantitative and qualitative responses [2]. Generalized linear model assumes a linear relationship between the mean of the dependent variable and the unknown parameter $\beta_1, \ldots, \beta_2$

$$\mu_{y|x} = E(y|x) = \beta_0 + \sum_{j=1}^{p} \beta_J X_j \quad (1)$$

or

$$g(\mu_{y|x}) = E(y|x) = \beta_0 + \sum_{j=1}^{p} \beta_J X_j \quad (2)$$

where $X_1,...X_p$ are independent variables [3], which replace the linear $\sum_{j=1}^{p} \beta_J X_j$ form by a sum of smooth functions $\sum_{j=1}^{p} f(x_j)$.

The multiple linear regression model:

$$y_i = \beta_0 + \beta_{1x_{i1}} + \beta_{2x_{i2}} + \cdots + \beta_{px_{ip}} + \varepsilon_i \quad (3)$$

needs to be extended in order to allow for non-linear relationships between each feature and the response. For this aim, each linear component $\beta_{ix_{ij}}$ is replaced with a (smooth) non-linear function $f_j(x_{ij})$. Then the model becomes:

$$y_i = \beta_0 + \sum_{j=1}^{p} f_j(x_{ij}) + \varepsilon_i \quad (4)$$

$$y_i = \beta_0 + f_1(x_{i1})\, f_2(x_{i2}) + \cdots + f_p(x_{ip}) + \varepsilon_i \quad (5)$$

This example of a GAM is called an additive model because we calculate a separate fj for each Xj and then add together all of their contributions [2].

In order to summarize the trend of a dependent variable Y as a function of one or more independent variables X1,...,Xp, a tool called smoother is used. It produces an estimate of the trend that variates less than Y itself. We call a smooth the estimate produced by a smoother. This is a very useful way to pick up the trend out of the plot. It also estimates the dependence of the mean of Y on the predictors.

Regression splines created by specifying a set of knots generate a sequence of basic functions. Then, least squares are used to estimate the spline coefficients. In fitting a smooth curve to a set of data, a g(x) function is determined. The latter fits the observed data well, on the condition that the residual sum of squares (RSS) be as low as possible:

$$RSS = \sum_{i=1}^{n} (y_i - g(x_i))^2 \quad (7)$$

where λ is a non-negative tuning parameter. The g function that minimizes (7) is referred to as a smoothing spline. This means that (7) takes the "Loss + Penalty" formulation that we encounter in the context of the Ridge regression and the Lasso regression. The term $\sum_{i=1}^{n} (y_i - g(x_i))^2$ is a loss function

that encourages g to fit the data well, and the $\lambda \int g''(t)^2 dt$ term is a penalty term that penalizes the variability in g and helps to avoid overly wiggly fits.

## 2. EXPERIMENTAL

The initial dataset consisted of the normalized ATR-FTIR spectra of 60 compounds of forensic interest. The spectra have been recorded in the 4000 - 400 cm-1 spectral window, 5 cm-1 apart, by calculating the average of 1868 scans. These spectra are publicly available, being provided by the Drug Enforcement Administration (DEA), USA. The samples belong to three sets of drugs: 2C-x and DOx psychedelic amphetamines (class 1), cannabinoids (class 2), and negatives (class 3), the latter being other substances of forensic interest that have been randomly selected. A second dataset, generated by selecting the most important 186 features by applying a Genetic Algorithm (GA), has been used for analysis and comparison [4].

## 3. RESULTS AND DISCUSSION

The model was developed in the R environment, by using the mgcv library. This library provides a generalized additive modelling function gam. The dataset was split into two subsets, for training and validating the model. The main goal was to maximize the efficiency of the system in predicting the class membership of the analyzed substances included in each test subset. The substances from the training and validating sets and their correct class membership are presented in Table 1 and Table 2.

Table 1. Substances included in the training set

| Training Set | | | | | |
|---|---|---|---|---|---|
| | Substance | Class | | Substance | Class |
| 1 | 25B-NBOMe HCl (Lot #N18-P1C) | 1 | 31 | AM2233 (Lot #0436043-18) | 2 |
| 2 | 25C-NB3OMe HCl (Lot #N17-P72C) | 1 | 32 | JWH-018 adamantyl-carboxamide | 2 |
| 3 | 25C-NB4OMe HCl (Lot #N17-P73C) | 1 | 33 | JWH-018 Benzimidazole | 2 |
| 4 | 25C-NBOMe HCl (Lot #N17-P71D) | 1 | 36 | JWH-019 (Lot #K8H81106) | 2 |
| 5 | 2,5-Dimethoxy-4-Chloro-amphetamine HCl | 1 | 37 | JWH-022 (Lot #ALB227-8) | 2 |
| 7 | 2,5-Dimethoxy-4-methylamphetamine HCl | 1 | 40 | JWH-122 (Lot #N1P3) | 2 |
| 11 | 25E-NBOMe HCl (Lot #N17-P97B) | 1 | 42 | JWH-203 (Lot #K8H81110) | 2 |
| 12 | 25H-NBOMe HCl (Lot #N16-P81A) | 1 | 43 | JWH-210 (Lot #N1P36EMG) | 2 |
| 13 | 25I-NB3OMe HCl (Lot #N17-P74D) | 1 | 45 | JWH-307 (Lot #0439287-1) | 2 |
| 14 | 25I-NB4OMe HCl (Lot #N17-P75C) | 1 | 46 | RCS-4 (Lot #N1P38EMG) | 2 |
| 15 | 25I-NBOMe Base (Lot #SF0003) | 1 | 47 | RCS-8 (Lot #ALB203-21) | 2 |
| 16 | 25I-NBOMe HCl (Lot #N17-P11B) | 1 | 48 | bk-MDDMA HCl (Lot #0432923-23) | 3 |
| 17 | 2C-B BZP diHCl (Lot #H-0416) | 1 | 49 | Bufotenine Oxalate Hydrate | 3 |
| 18 | 2C-B HCl (Lot #729.1B4.1) | 1 | 50 | Buphedrone HCl | 3 |
| 19 | 2C-E HCl (Lot #H-0407) | 1 | 52 | BZP diHCl (Lot #ALB90-5) | 3 |
| 20 | 2C-I HCl (Lot #2TDM-37-04) | 1 | 53 | Cathine Base (Lot #N11-P-17) | 3 |
| 21 | 2C-T-2 HCl (Lot #N1P31) | 1 | 55 | Cathinone HCl (Lot #113-1191-12) | 3 |
| 23 | 3,4-Dimethoxyamphetamine HCl | 1 | 56 | CB-13 (Lot #N1P15) | 3 |
| 24 | 5-Methoxy-alpha-methyltryptamine HCl | 1 | 58 | CP 47,497 (Lot #0419860-11) | 3 |
| 26 | 5-Methoxy-N,N-diethyltryptamine HCl | 1 | 60 | Diisopropyltryptamine HCl | 3 |

Table 2. Substances included in the validation set

| Validation Set | | |
|---|---|---|
| | Substance | Class |
| 6 | 2,5-Dimethoxy-4-ethylamphetamine HCl (Lot #J-1) | 1 |
| 8 | 2,5-Dimethoxyamphetamine HCl (Lot #AKB29A) | 1 |
| 9 | 2,5-Dimethoxyphenethylamine HCl (Lot #MP137-139) | 1 |
| 10 | 25D-NBOMe HCl (Lot #N17-P88C) | 1 |
| 22 | 2C-T-7 HCl (Lot #2TDM-198-01) | 1 |
| 25 | 5-Methoxy-Diallyltryptamine HCl (Lot #RM-131001-04) | 1 |
| 27 | d,l-DOB HCl (d,l-4-Bromo-2,5-dimethoxyamphetamine HCl) | 1 |
| 28 | Methoxetamine HCl (Lot #N16-P100C) | 1 |
| 34 | JWH-018 (Lot #ALB045RC_183-1) | 2 |
| 35 | JWH-018 N-(5-chloropentyl) analog (Lot #0434099-14) | 2 |
| 38 | JWH-073 (Lot #0409793-37) | 2 |
| 39 | JWH-081 (Lot #ALB056RC) | 2 |
| 41 | JWH-200 (Lot #0424688-3) | 2 |
| 44 | JWH-250 (Lot #ALB055RC) | 2 |
| 51 | Butylone HCl (Lot #2011DEA003-25A) | 3 |
| 54 | Cathine HCl (Lot #284) | 3 |
| 57 | CP 47,497 C8 homologue (Lot #0425163-4) | 3 |
| 59 | dihydro-PPP HCl (Lot #RM-131218-03) | 3 |

First, the full dataset was used to train the model. The classification of substances included in the validation subset was done by using the values of fitted values, which led to the following confusion matrix:

| | *Reference* |
|---|---|
| *Prediction* | 1 2 3 |
| | 1 4 1 0 |
| | 2 4 4 2 |
| | 3 0 1 2 |

The class membership predicted by the system is presented in Table 3 and Figure 1.

Table 3. Validation results for the system trained with the full dataset

| Substance | Predicted class | Class |
|---|---|---|
| 6 | 1.554746 | 1 |
| 8 | 1.428204 | 1 |
| 9 | 1.73939 | 1 |
| 10 | 1.07665 | 1 |
| 22 | 1.664468 | 1 |
| 25 | 1.206176 | 1 |
| 27 | 1.604426 | 1 |
| 28 | 1.431155 | 1 |
| 34 | 2.42349 | 2 |
| 35 | 1.261547 | 2 |
| 38 | 2.3573 | 2 |
| 39 | 2.177051 | 2 |
| 41 | 2.319647 | 2 |
| 44 | 2.76023 | 2 |
| 51 | 2.198402 | 3 |
| 54 | 2.01953 | 3 |
| 57 | 2.837992 | 3 |
| 59 | 3.406174 | 3 |

We can observe that the classification performance is low in the case of the compounds included in class 1, i.e. the 2C-x and DOx psychedelic amphetamines. In their case, the overall accuracy is 0.4444.

The variation of the normalized error of training vs. the number of learning cycles is presented in Figure 2 for the 40MP&C-ANN network. The relative importance, determined for the first 40 most important descriptors of the 40MP&C-ANN system built with 40 molecular properties and charge descriptors is shown in Figure 3 in descending order.

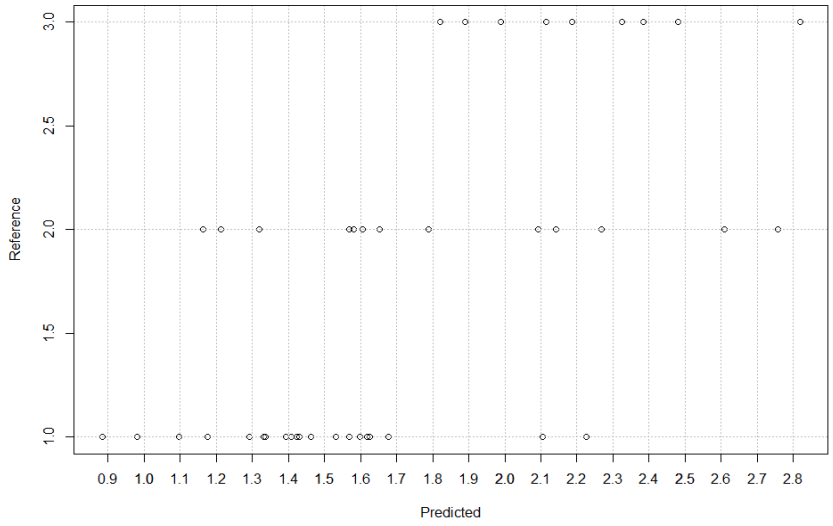


*Fig. 1. Classification results obtained with the system trained by using the full dataset.*

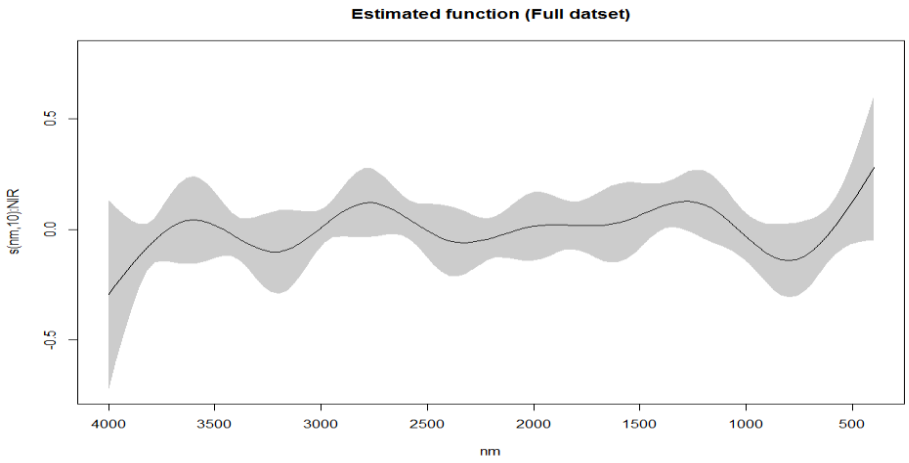


*Fig. 2. Estimated function obtained for the system trained with the full dataset.*

The model built for the GA selected dataset leads to a much better overall accuracy, i.e. 0.7778, with the following confusion matrix:

| | *Reference* | | |
|---|---|---|---|
| *Prediction* | 1 | 2 | 3 |
| 1 | 7 | 1 | 0 |
| 2 | 1 | 4 | 1 |
| 3 | 0 | 1 | 3 |

The predicted class memberships obtained for the classification of the substances included in the validation subset for the system trained with the GA selected dataset are presented in Table 4.

Table 4. Validation results for the system trained with the GA selected dataset

| Substance | Predicted class | Class |
|---|---|---|
| 6 | 0.945814 | 1 |
| 8 | 1.451981 | 1 |
| 9 | 1.4488305 | 1 |
| 10 | 1.2767805 | 1 |
| 22 | 0.9613652 | 1 |
| 25 | 1.165072 | 1 |
| 27 | 1.5285436 | 1 |
| 28 | 1.14741 | 1 |
| 34 | 2.2115683 | 2 |
| 35 | 1.5355147 | 2 |
| 38 | 1.2372119 | 2 |
| 39 | 2.1707806 | 2 |
| 41 | 1.8426685 | 2 |
| 44 | 2.6425833 | 2 |
| 51 | 2.9148402 | 3 |
| 54 | 2.1443821 | 3 |
| 57 | 3 | 3 |
| 59 | 3 | 3 |

The estimated smooth functions computed for both datasets by using adaptive smother are presented in Fig. 2 and Fig. 3. When the irrelevant features are removed from the spectra by using GA, the model performs much better in classifying the substances from Class 1. Overall, a number of 7 substances are misclassified by the system trained with the GA selected dataset.

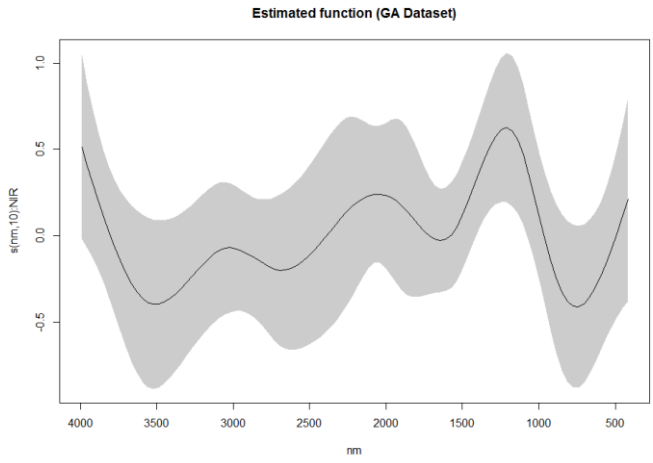


*Fig. 3. Estimated function obtained for system trained with the GA selected dataset.*

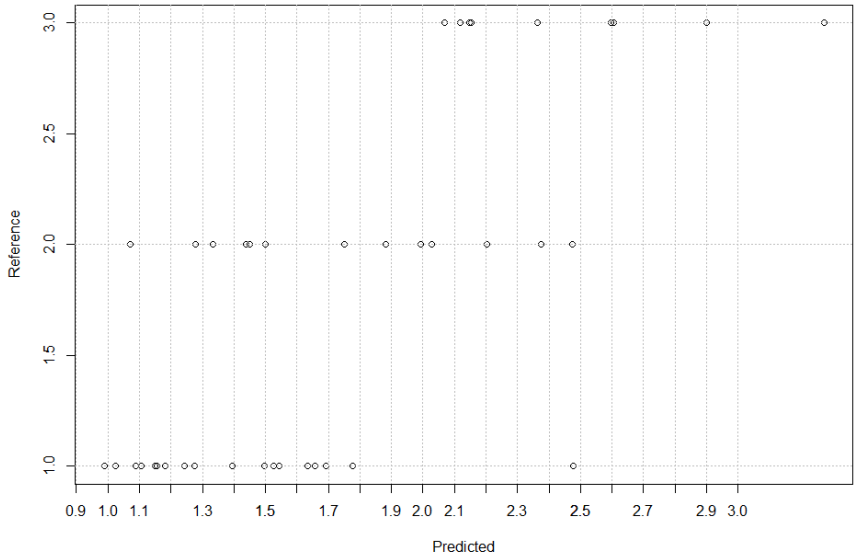


*Fig. 4. Classification results obtained with the system trained with the GA selected dataset.*

## 4. CONCLUSIONS

Two models have been compared from the point of view of their performances is assigning the class identity based on ATR-FTIR full and GA selected spectra. The results indicate that the classification performance of the CAM model increases when the system is trained with a dataset containing only the relevant spectral features, the latter being selected by applying the GA.

**Acknowledgement:** We acknowledge the funding from ENPI project code MIS ETC 1676 INPOLDE (2013-2015), Joint Operational Programme Romania-Ukraine-Republic of Moldova 2007-2013.

## References

1. European Monitoring Centre for Drugs and Drug Addiction, European Drug Report 2018 - Trends and Developments, Luxembourg: Publications Office of the European Union, 2018.
2. J. Witten, D. Hastie, and R. Tibshirani, "An introduction to statistical learning", vol. 112, p. 18. New York: Springer, 2013.
3. T. Hastie and R.. Tibshirani. "Generalized additive models", vol. 43, CRC Press, 1990.
4. C. Negoita, M. Praisler, and A. Ion, "Artificial Intelligence Application Designed to Screen for New Psychoactive Drugs Based on their ATRFTIR spectra", AIP Conference Proceedings 2075, 170026, 2019.
5. S. Wood, Generalized Additive Model: an introduction with R, Chapman and Hall/CRC.