ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
MATHEMATICS, PHYSICS, THEORETICAL MECHANICS
FASCICLE II, YEAR XI (XLII) 2019, No. 1

Article DOI: https://doi.org/10.35219/ann-ugal-math-phys-mec.2019.1.08

PATTERN RECOGNITION ANALYSIS OF THE CLASS IDENTITY RECOGNITION EFFICIENCY OF A PORTABLE LASER INFRARED SENSOR DETECTING AMPHETAMINES AND THEIR MAIN PRECURSORS

Mirela Praisler[1], Stefanut Ciochina[2]

[1]Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment, „Dunărea de Jos" University of Galati, Romania, e-mail: Mirela.Praisler@ugal.ro
[2]Faculty of Sciences and Environment, Department of Mathematics and Computer Science, „Dunărea de Jos" University of Galati, Romania, e-mail: Stefanut.Ciochina@ugal.ro

Abstract
We are presenting a pattern recognition analysis assessing the class identity recognition efficiency of a portable laser infrared sensor detecting controlled phenethylamines, i.e. the stimulant and hallucinogenic amphetamines, as well as ephedrines, which are their main precursors. The training set consists of laser infrared spectra of the later compounds and of negatives, which are randomly selected non-amphetamines. The spectra have been recorded in the spectral domain 1405 - 1150 cm^{-1}, preprocessed with a w_{TE}^2 Fisher discriminating function, and then subjected to Principal Component Analysis (PCA).
The PCA scores have been used in order to build several pattern recognition systems designed to recognize the class identity of the targeted compounds, i.e. Cluster Analysis and Naive Bayesian Classifier. The detection efficiency obtained for these two systems is presented and discussed in detail.

Keywords: Amphetamines and precursors, Cluster Analysis, Naive Bayesian Classifier

1. INTRODUCTION

Drug abuse is an increasingly problem for the public health systems, which is also incurring important social and economic costs worldwide [1-3]. Phenetylamines, i.e. amphetamines and their main precursors – the ephedrines, are some of the most popular designer drugs: the proliferation of clandestine laboratories has made these synthetic drugs increasingly afordable [4]. Statistics show that the number of these laboratories has increased especially in Europe [5]. One of the main measures taken during the last decade by the European Community, in order to fight against narcotraffic, was to encourage the development of new portable analytical instruments, able to detect efficiently illegal phenetylamines as well as their main precursors [6].

In this paper we are presenting an artificial intelligence application that was developped in order to operate a new portable laser GC-IRAS spectrometer, which was developped within the DIRAC EU project [7, 8]. The application was designed to provide: a) a fast automatic screening for amphetamines and for ephedrines; b) a way to predict the pharmacological activity of the detected amphetamine

(stimulant or hallucinogenic); c) a user-friendly human-machine interface, making the instrument easy to use for law enforcement officers without needing a strong background in analytical chemistry.

2. EXPERIMENTAL PART

The database consists of the infrared spectra of 36 compounds, recorded between 1405 and 1150 cm^{-1}, which is the domain where the UT8 quantum cascade laser (QCL) that equipps the portable spectrometer is emitting. The absorption has been measured every 5 cm^{-1} [9, 10]. The spectral input consists of 19 positives and 17 negatives (non-phenethylamines, class code N). The positives are 7 stimulant amphetamines (amphetamine, analogues and homologues, class named M), 6 ephedrines (ephedrine, stereoisomers and diastereomers, class named E), 6 hallucinogenic amphetamines (3,4-methylenedioxyamphetamines and analogues, class named T) [11-13].

A w_{TE} Fisher discriminating function was computed based on the spectra included in the spectral database [14]. For this purpose, the positives have been divided into two classes: class I includes the spectra of the T and E drugs of abuse, while class II consists of the remaining spectra (M and N compounds) [15-17]. The mean spectra of the original and of the w_{TE}^2 processed spectra included in class I are presented in Fig. 1, while those included in class II are illustrated in Fig. 2. We may notice that the main effect of the w_{TE}^2 spectra processing is the enhancement of the intensity differences of the absorptions that the targetted compunds have around 1245 and 1190 cm^{-1}. More specifically, the positives (T and E) included in class I are characterised by much (more than 3 times) stronger absorptions at these wavenumbers than the compounds (M and N) included in class II.

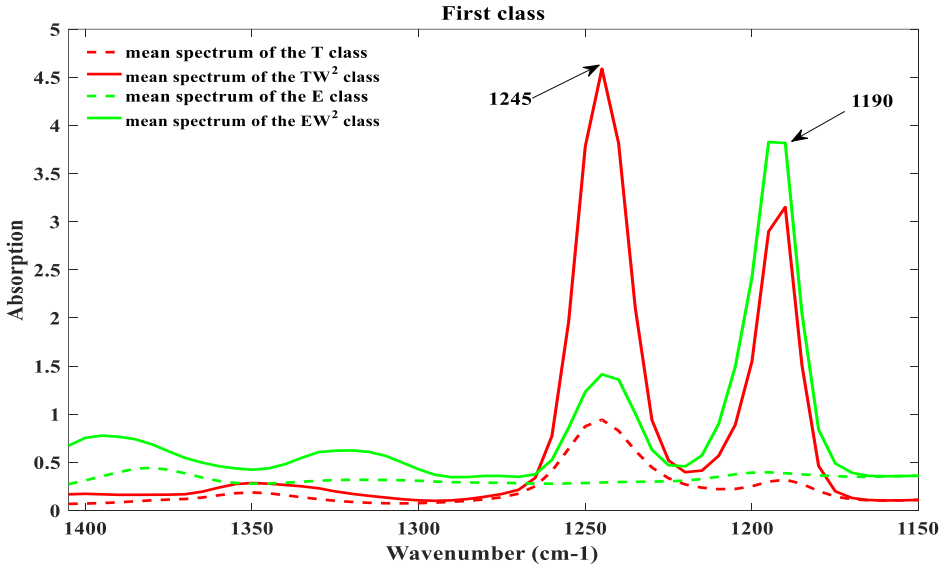

Fig. 1. Mean spectra of the original and w_{TE}^2 processed spectra of the hallucinogenic amphetamines (class name T) and of the ephedrines (class name E) included in the database used for multivariate analysis

The w_{TE}^2 processed spectra have been assessed by Principal Component Analysis (PCA) [18], by using the MATLAB software. The number of principal components (PCs) have been established based on the cumulated explained variance. The resulting 3D score plot has been used in order to determine the clusters that may be reliably distinguished and used for the discrimination of the modeled drugs of abuse.

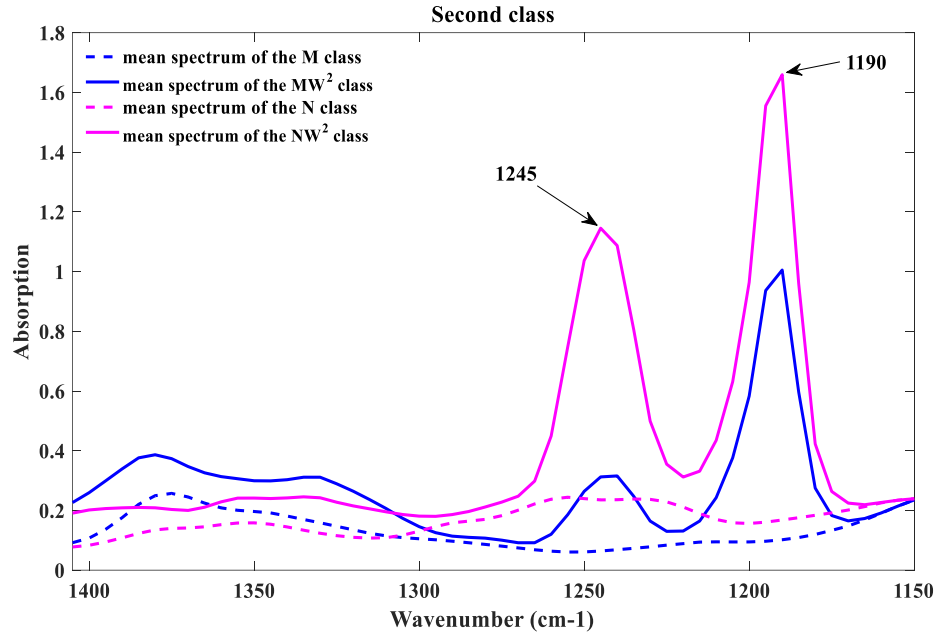

Fig. 2. Mean spectra of the original and w_{TE}^2 processed spectra of the stimulant amphetamines (class name M) and of the negatives (class name N) included in the database used for multivariate analysis

The visual inspection of the 3D score plot has been corroborated with the results of a Hierarchical Cluster Analysis (HCA) [19, 20]. The first assessed dendrogram has been built with the first two PCs. The Silhouette index [21] was used to assess the number of clusters that can be reliably detected with this clustering tree. The results have been compared with those obtained with the first three PCs. Finally, these results have been compared with those yielded by the Naïve Bayes classifier [22].

3. RESULTS AND DISCUSSION

The w_{TE}^2 processed spectra have been evaluated by Principal Component Analysis (PCA), in order to assess to what extent the targetted classes of designer drugs and precursors are forming clearly distinguishable clusters. The dynamics of the explained variance is presented in Table 1. It indicates that the first two PCs are enough to obtain a cumulated explained variance larger than 90%. If the first three PCs are taken into account, a cumulated explained variance of 95.0589% is obtained.

Table 1. Reducing the dimensionality of the hyperspace formed by the w_{TE}^2 preprocessed spectra

Principal Component	Explained variance (%)	Cumulated explained variance (%)
PC1	72.0021	72.0021
PC2	19.9755	91.9776
PC3	3.0813	95.0589

Hence, PCA was performed with the first three PCs. The 3D score plot obtained with these PCs is presented in Fig. 3. It indicates that all three classes of designer drugs (M, T) and precursors (E) are

clustering. Among the positives, the hallucinogenic amphetamines are forming the densest cluster and the ephedrines (E) form the most dispersed cluster. However, the cluster formed by stimulant amphetamines (M) is closely surrounded by the cloud formed by the negatives, making it difficult to clearly distinguish the class identity of an unknown whose associated point is located towards the periphery of the M cluster. Although their number is much smaller than in the case of the M amphetamines, some negatives are also located rather close to the cluster formed by the ephedrines (E).

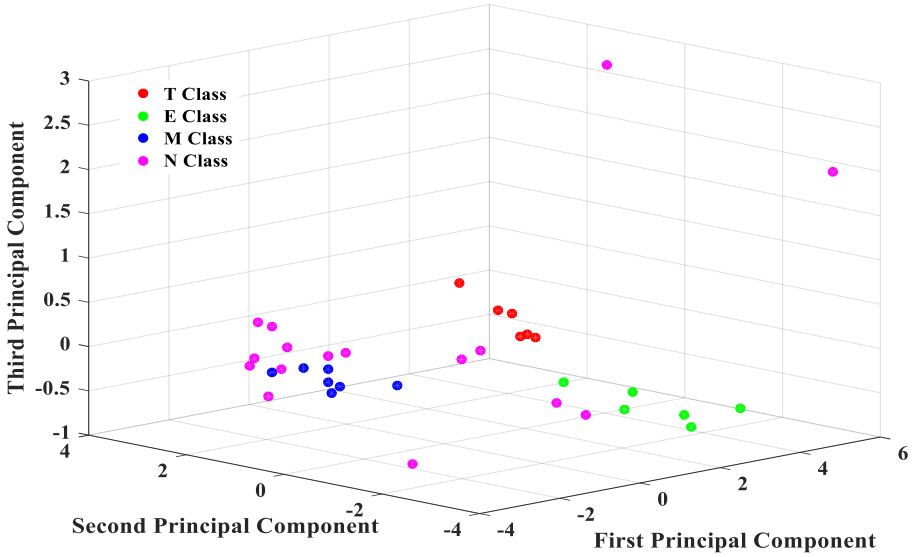

Fig. 3. 3D score plot obtained with the w_{TE}^2 processed spectra based on Principal Component Analysis (PCA) performed with the first three principal components (PC1, PC2, PC3)

As this qualitative analysis could not provide a clear answer regarding the number of clusters that may be distinguished reliably, a quantitative assessment was performed based on the HCA of the (PC1, PC2, PC3) scores obtained for the w_{TE}^2 processed spectra. The Silhouette index has been determined and its dynamics is presented in Fig. 4. It clearly indicates that a number of 3 clusters may be distinguished reliably. Hence, two of the modeled classes of compounds cannot be discriminated.

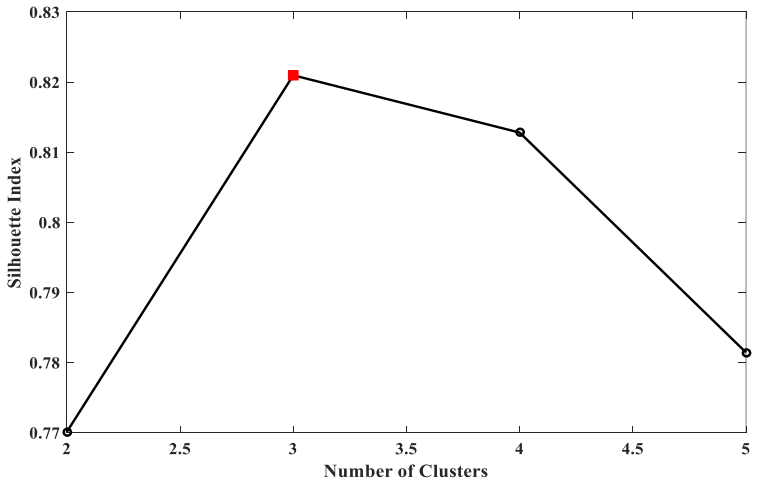

Fig. 4. Dynamics of the Silhouette index determined based on the (PC1, PC2, PC3) scores obtained with the w_{TE}^2 processed spectra

The associated clustering tree, presented in Fig. 5, shows that, in the case of this technique, the cluster formed by the hallucinogenic amphetamines (T) is affected by (only) one false positive (N28). No T compound shows in the other clusters, so there are no T false negatives. The ephedrines (E) are forming an even better defined cluster: There are no false positives, nor E false negatives. On the other hand, there is a significant overlap between the cluster of the stimulant amphetamines (M) and the cluster formed by the negatives (N). Hence, the three clusters that may be reliably distinguished based on this dendrogram are the T, E and (M, N) clusters.

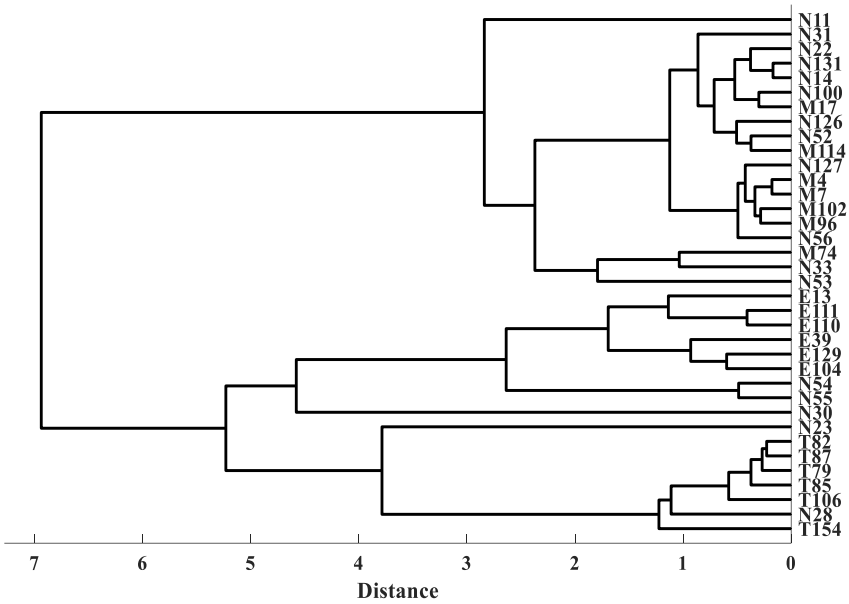

Fig. 5. Clustering tree determined based on the (PC1, PC2, PC3) scores obtained with the w_{TE}^2 processed spectra

In these conditions, it was worth testing if the same results may be obtained with the scores of only the first two PCs, i.e. PC1 and PC2, which are cumulating an explained variance of 91.9776%, i.e. only 3.0813% less than the 95.0589% explained variance cumulated by the first three PCs.

The dynamics of the Silhouette index determined with the (PC1, PC2) scores is presented in Fig. 6. It indicates that the same number (three) of clusters may be clearly distinguished with only the first two PCs. It is worth underlining that the value of the Silhouette index determined for 3 clusters with the (PC1, PC2) scores is even (slightly) larger than the value obtained for (PC1, PC2, PC3) scores (see Fig. 4 and Fig. 6). Hence, the three clusters may be even better distinguished with only the first two PCs.

The associated dendrogram, presented in Fig. 7, shows that this clustering tree ensures the same sensitivity and selectivity than the clustering tree built with the (PC1, PC2, PC3) scores. Hence, using the (PC1, PC2) dendrogram is the simplest and fastest way of detecting T and E drugs.

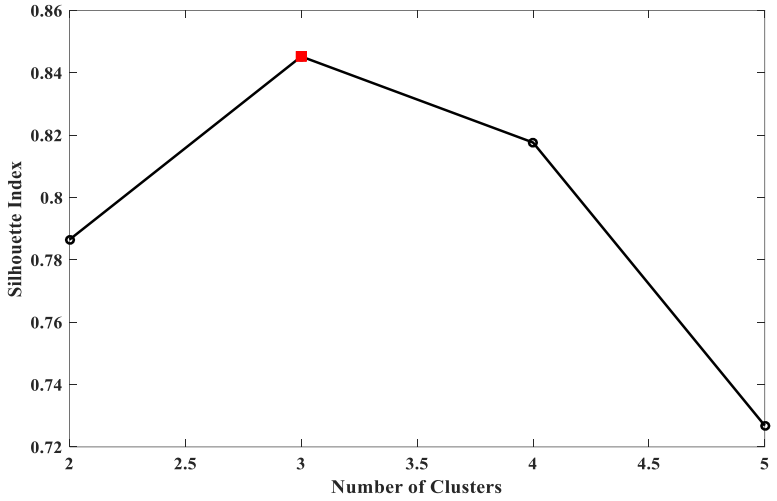

Fig. 6. Dynamics of the Silhouette index determined based on the (PC1, PC2) scores obtained with the w_{TE}^2 processed spectra

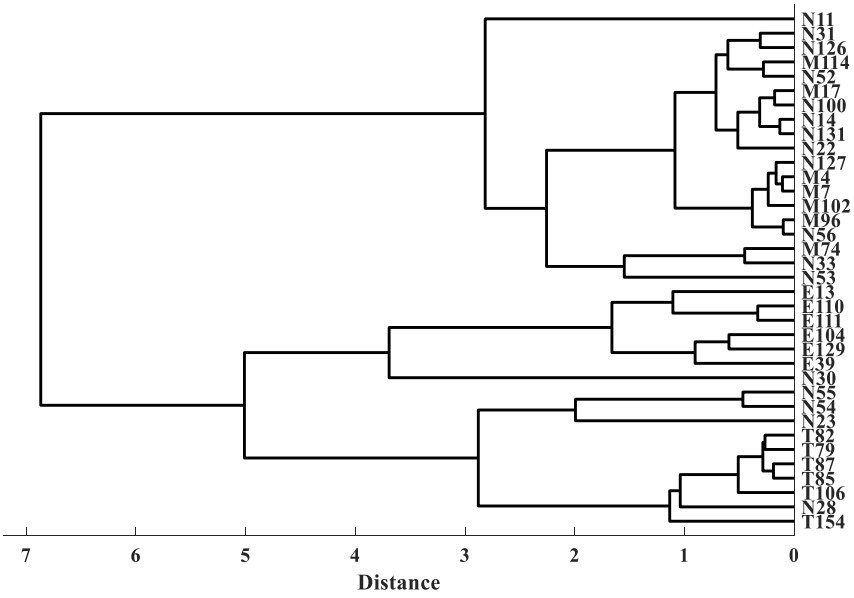

Fig. 7. Clustering tree determined based on the (PC1, PC2) scores obtained with the w_{TE}^2 processed spectra

These results have been compared with those yielded by the Naïve Bayes classifier. Table 2 presents the misclassifications yielded by this classifier. By comparing these results with the clustering tree presented in Fig. 7, we may draw the conclusion that both methods misclassify practically the same compounds. This means that the spectra, recorded in the narrow spectral window of the laser source that equipps the portable laser GC-IRAS spectrometer (1405 - 1150 cm^{-1}) of these compounds does not contain enough information about the specificity of these molecular structures.

Table 2. Misclassifications yielded by the Naïve Bayes classifier

Tested	Answer	Posterior				Cost			
		T	E	M	N	T	E	M	N
M74	N	0	0	0.4018	**0.5981**	1	1	0.5981	**0.4018**
N56	M	0	0	**0.9102**	0.0897	1	1	**0.0897**	0.9102
N52	M	0	0	**0.7984**	0.2015	1	1	**0.2015**	0.7984
N126	M	0	0	**0.7114**	0.2885	1	1	**0.2885**	0.7114
N100	M	0	0	**0.5040**	0.4959	1	1	**0.4959**	0.5040
N127	M	0	0	**0.8949**	0.1050	1	1	**0.1050**	0.8949

4. CONCLUSIONS

PCA has allowed a qualitative analysis of the possibility of detecting any new compound with a molecular structure similar to stimulant and hallucinogenic amphetamines, as well as to ephedrine. The results have been corroborated with a quantitative analysis based on HCA and the Silhouette index, which firmly indicated that only hallucinogenic amphetamines and ephedrines may be reliably detected based on the PCA scores determined with their w_{TE}^2 processed spectra recorded in the 1405 - 1150 cm^{-1} spectral window. The results have also indicated that the scores of the first two PCs are enough for building a clustering tree that may reliably detect hallucinogenic amphetamines (T) and ephedrines (M). The use of the Naïve Bayes classifier does not improve significantly the number of misclassifications.

Acknowledgements

Part of the research has been funded by EC under the grant agreement n° FP7-SEC-2009-242309 DIRAC. The work of Stefanut Ciochina has been funded by the Romanian Ministry of European Funds within the POSDRU/107/1.5/S/76822 project. The authors are grateful for the financial support.

References

1. United Nations Office on Drugs and Crime (UNODC), *World Drug Report 2018. Executive Summary Conclusions and Policy Implications*, United Nations, Vienna, 2018.
2. United Nations Office on Drugs and Crime (UNODC), *World Drug Report 2018. Drugs and Age: Drugs and Associated Issues Among Young People and Older People*, United Nations, Vienna, 2018.
3. United Nations Office on Drugs and Crime (UNODC), *World Drug Report 2018. Women and Drugs: Drug Use, Drug Supply and their Consequences*, United Nations, Vienna, 2018.
4. United Nations Office on Drugs and Crime (UNODC), *World Drug Report 2018. Global Overview of Drug Demand and Supply: Latest Trends, Cross-Cutting Issues*, United Nations, Vienna, 2018.
5. United Nations Office on Drugs and Crime (UNODC), *World Drug Report 2018. Analysis of Drug Markets Opiates: Cocaine, Cannabis, Synthetic Drugs*, United Nations, Vienna, 2018.
6. European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), *European Drug Report 2018: Trends and Developments*, Publications Office of the European Union, Luxembourg, 2018.

7. Herbst J., Scherer B., Ruf A., Erb J., Lambrecht A., Hollow fiber based quantum cascade laser spectrometer for fast and sensitive drug identification, *The International Society for Optical Engineering* SPIE 8268, 82681Q (2012).
8. Mengali S., Liberatore N., Luciani D., Viola R., Cardinali G. C., Elmi I., Poggi A., Zampolli S., Biavardi E., Dalcanale E., Bonadio F., Delemont O., Esseiva P., Romolo F. S., Rapid screening and identification of illicit drugs by IR absorption spectroscopy and gas chromatography, *The International Society for Optical Engineering* SPIE 8631, Quantum Sensing and Nanophotonic Devices X, 86312F (2013).
9. Praisler M., Ciochina S., Negoita C., Improved Selectivity in Detecting Controlled Amphetamines and their Main Precursors based on Laser Infrared Spectra, *2017 E-Health and Bioengineering Conference, EHB 2017*, 28 July 2017, pp. 233-236. Article number 7995404.
10. Ciochina S., Praisler M., Negoita C., Cluster Analysis Evaluating the Automated Detection of Drugs of Abuse with a New Hollow Fiber based Quantum Cascade Laser Infrared Spectrometer, *2017 E-Health and Bioengineering Conference, EHB 2017*, 28 July 2017, pp. 237-240. Article number 7995405.
11. Praisler M., Ciochina S., Coman M., Screening for Illicit Psychoactive Drugs Based on Pattern Recognition Methods, *5th International Symposium on Electrical and Electronics Engineering, ISEEE 2017*, 20-22 October 2017, Galati, Romania
12. Ciochina S., Praisler M., Coman M., Hierarchical Cluster Analysis Applied for the Automated Recognition of Psychoactive Substances and of Their Main Precursors, *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, 20-22 October 2017, Galati, Romania.
13. Praisler M., Ciochina S., Coman M., Hunting for Illicit Psychoactive Substances and Precursors: a Multivariate Approach, *2017 21st International Conference On System Theory, Control And Computing (ICSTCC)*, pp 248-253. Article number 8107042.
14. Ciochina S., Praisler M., Detecting Phenethylamines According To Their Pharmacological Activity, Annals Dunarea de Jos Univ. Galati, Fascicle II, Year XLI, 18, 2018.
15. Gosav S., Praisler M., Dorohoi D. O., Popa G., Structure – Activity Correlations for Illicit Amphetamines Using ANN and Constitutional Descriptors, Talanta 70, 922, 2006.
16. Gosav S., Praisler M., Dorohoi D. O., ANN Expert System Screening for Illicit Amphetamines using Molecular Descriptors, J. Mol. Struct. 834-836, 188, 2007.
17. Gosav S., Praisler M., Artificial Neural Networks Built for the Recognition of Illicit Amphetamines Using a Concatenated Database, Rom. Rep. Phys. 54 (9/10), 929, 2009.
18. Jolliffe I. T., *Principal Component Analysis*, 2nd ed., Springer, New York, 2002.
19. Mullner D., *Modern hierarchical aglomerative clustering algoritms*, arXiv:1109.2378v1, 2011.
20. Kaufman L., Rousseeuw P. J., *Finding Groups in Data. An Introduction to Cluster Analysis*, Willey, New Jersey, 2005.
21. Starczewski A., Krzyżak A., Performance Evaluation of the Silhouette Index, in Rutkowski L. et al. (Eds), *Artificial Intelligence and Soft Computing ICAISC 2015, Lecture Notes in Computer Science, vol. 9120*, Springer, Switzerland, 2015, pp. 49-58.
22. Cichosz P., Naïve Bayes classifier, in Cichosz P., *Data Mining Algorithms: Explained Using R*, John Wiley & Sons, 2015, pp.118-133.