

Research Article

A QSRR Modeling of Hazardous Psychoactive Designer Drugs Using GA-PLS and L-M ANN

Hamzeh Karimi,¹ Hadi Noorizadeh,² and Abbas Farmany²

¹ Faculty of Sciences, Islamic Azad University, South Tehran Branch, Tehran, Iran

² Faculty of Science, Islamic Azad University, Ilam Branch, Ilam, Iran

Correspondence should be addressed to Hamzeh Karimi, h.karimi744@yahoo.com

Received 29 January 2012; Accepted 5 March 2012

Academic Editors: I. Brondz and D. Gavril

Copyright © 2012 Hamzeh Karimi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The hazardous psychoactive designer drugs are compounds in which part of the molecular structure of a stimulant or narcotic has been modified. A quantitative structure-retention relationship (QSRR) study based on a Levenberg-Marquardt artificial neural network (L-M ANN) was carried out for the prediction of the capacity factor (k') of hazardous psychoactive designer drugs that contain Tryptamine, Phenylethylamine and Piperazine. The genetic algorithm-partial least squares (GA-PLS) method was used as a variable selection tool. A PLS method was used to select the best descriptors and the selected descriptors were used as input neurons in neural network model. For choosing the best predictive model from among comparable models, square correlation coefficient (R^2) for the whole set is suggested to be a good criterion. Finally, to improve the results, structure-retention relationships were followed by nonlinear approach using artificial neural networks and consequently better results were obtained. Also this demonstrates the advantages of L-M ANN. This is the first research on the QSRR of the designer drugs using the GA-PLS and L-M ANN.

1. Introduction

Designer drugs (sometimes also referred to as club drugs) are a particular class of synthetic drugs most often associated with underground youth dance parties called raves, wherein participants listen to techno music and experiment with psychoactive substances. These drugs have been created by changing the molecular structure of other existing drugs, to create something new with similar pharmacological effects, hence, the name designer drug. They are plentiful, cheap, and dangerous. For example, the pharmaceutical drug amphetamine (which was originally created as an anesthetic) has been modified to be 80 to 1,000 times more potent than heroin. Prepared by underground, amateur chemists known as cooks, designer drugs can be injected, smoked, snorted, or ingested. These synthetic drugs can be easily obtained on the street or on the Internet.

Once changed, they become known by a variety of street names, for example, XTC, Ecstasy, Adam, Lover's Speed, Special K, Fantasy, and Nature's Quaalude. Most have a

rapid onset of affect (1 to 4 minutes) and a short duration of action (generally 30–90 minutes, and no more than a few hours). They are sold as tablets or capsules, and often produce feelings of stimulation and euphoria, a sense of well-being, and various sensory distortions. Higher doses can lead to paranoia, hallucinations, violent or otherwise irrational behavior, and fatal overdosing. Some designer drugs are depressants, so they are used when an individual is coming down from a stimulant like Ecstasy.

In general, the physical symptoms common among users of psychoactive designer drugs include hypertension, increased heart rate, clenched teeth, blurred vision, uncontrolled tremors, anorexia, nausea and vomiting, impaired speech, seizures, permanent brain damage, and death.

Some common psychological side effects include confusion, irritability, severe anxiety, extreme emotional sensitivity, depression, amnesia, violent behavior, insomnia, and hallucinations [1–4].

As shown in Figure 1, these hazardous psychoactive designer drugs have I, II, and III basic skeletons. Type I

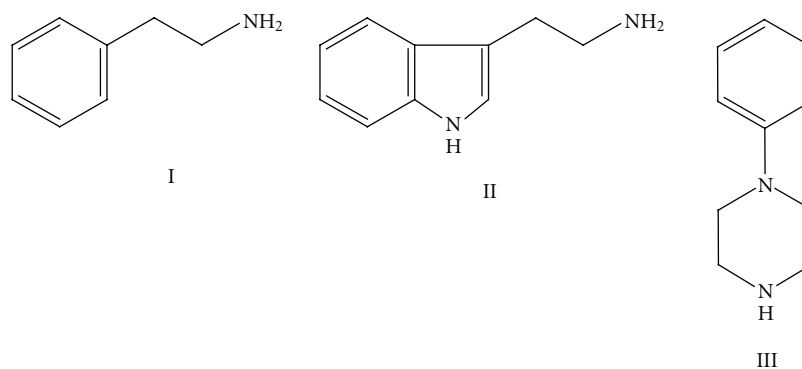


FIGURE 1: Structures of basic skeletons. I: phenylethylamine; II: tryptamine; III: phenylpiperazine.

structure is phenylethylamine-(PEA-) related compounds with structural similarities to both amphetamine and the psychedelic PEA, mescaline. Type II structure is tryptamines-(T-) related compounds with structural similarities to hallucinogenic psilocin. Type III structure is phenylpiperazine-(PP-) related compounds with structural similarities to stimulus effects 1-(3-trifluoromethylphenyl)piperazine (TFMPP).

Most of the best known research chemicals are structural analogues of tryptamines or phenethylamines, but there are also many other completely unrelated chemicals which can be considered as part of the group. It is very difficult to determine psychoactivity or other pharmaceutical properties of these compounds based strictly upon structural examination. Many of the substances have common effects whilst structurally different and vice versa. As a result of no real official naming for some of these compounds, as well as regional naming, this can all lead to (and is anecdotally known to have led to) potentially hazardous mix-ups for users.

In order to prevent damage resulting from drug abuse, it is necessary to analyze the active ingredients, publicize the risks of these compounds, and, if illegal, quickly act to regulate them. For that, the library required the screening of these compounds, while there were a few of data for gas chromatography-mass spectrometry (GC/MS) [5–7] and liquid chromatography-mass spectrometry (LC/MS) [8, 9] and there was little data on hazardous psychoactive designer drugs. Also, there were few library systems for the liquid chromatography with photodiode array spectrophotometry (LC/PDA) [10, 11]. In order to eliminate the impact of the dead volume of the chromatographic system, the capacity factor (k') was calculated according to

$$k' = \frac{(t_r - t_0)}{t_0}, \quad (1)$$

where t_r is the retention time of the designer drugs, and t_0 is column void volume time of a nonretained compound or the dead time. It is known that the capacity factor (k') of a substance is related to the partition process, adsorption process, or both.

Using chemometrics tools to predict drugs and chemical tissue distribution, membrane permeability or biphasic

system partition is of major importance in physicochemical, environmental, and life sciences. Chemical distribution phenomena depend not only on molecular structure but also on the properties of the system in question [12]. Quantitative structure-retention relationship (QSRR) techniques based on different molecular descriptors have been successfully used to model organic chemicals properties [13]. A number of reports, deal with QSRR calculation of several compounds, have been published in the literature [14–16]. The QSRR models apply to partial least squares (PLSs) methods often combined with genetic algorithms (GAs) for feature selection [17–19]. Because of the complexity of relationships between the property of molecules and structures, nonlinear models are also used to model the structure-property relationships. Levenberg-Marquardt artificial neural network (L-M ANN) is nonparametric nonlinear modeling technique that has attracted increasing interest. In the present study, GA-PLS and L-M ANN were employed to generate QSRR models that correlate the structure of hazardous psychoactive designer drugs, with observed k' . This is the first research on the QSRR of the designer drugs using the GA-PLS and L-M ANN.

2. Computational

2.1. Data Set. Capacity factor (k') of 104 hazardous psychoactive designer drugs taken from the literature [20] is presented in Table 1. There are 51 types of PEA compounds with a type I structure, where these have a PEA skeleton such as 3,4,5-trimethoxyamphetamine (TMA), 2,5-dimethoxy-4-ethylthiophenethylamine (2C-T-2), 2,5-dimethoxy-4-propylthiophenethylamine (2C-T-7), and 4-bromo-2,5-dimethoxyphenethylamine (2C-B). There are 32 types of T compounds with a type II structure, where these have a T skeleton such as bis(methylethyl)[2-(5-methoxyindol-3-yl)ethyl]amine (5-MeO-DIPT) and 1-indol-3-ylprop-2-ylamine (AMT), and there are 21 types of PP compounds with a type III structure, where these have a piperazine skeleton such as 1-(3-chlorophenyl)piperazine (3CPP), 1-(4-methoxyphenyl)piperazine (4MPP), and TFMPP. These are tested using LC/PDA and GC/electron ionization (EI)/MS, and a library is created based on the analysis data obtained. The data registered into the library consisted of the capacity

TABLE 1: The data set, structure, and the corresponding observed k' values.

No.	Name	Structure	k'_{EXP}
Calibration set			
1	N-[2-(5-Methoxyindol-3-yl)ethyl]acetamide (Melatonin)	$\text{C}_{13}\text{H}_{16}\text{N}_2\text{O}_2$	0.09
2	2-Amino-3-(5-hydroxyindolin-3-yl)propanoic acid,5-hydroxy-tryptophan (5-HO-TP)	$\text{C}_{11}\text{H}_{12}\text{N}_2\text{O}_3$	0.31
3	3-Methoxy-tyramine (3-MT)	$\text{C}_9\text{H}_{13}\text{NO}_2$	0.39
4	3-(2-Aminoethyl)indol-5-ol (5-HO-T)	$\text{C}_{10}\text{H}_{12}\text{N}_2\text{O}$	0.46
5	(Dimethylaminoethyl)-1H-indol-5-ol (Bufotenin)	$\text{C}_{12}\text{H}_{16}\text{N}_2\text{O}$	0.50
6	p-Methoxyphen ethylamine	$\text{C}_9\text{H}_{13}\text{NO}$	0.53
7	2-(3,4-Dimethoxyphenyl)ethylamine (DMPEA)	$\text{C}_{10}\text{H}_{15}\text{NO}_2$	0.59
8	1-Methyl-3,4-dihydrobeta-carbolin-7-ol (Hannalol)	$\text{C}_{12}\text{H}_{12}\text{N}_2\text{O}$	0.68
9	2-(3,4-Dimethoxyphenyl)-N-methylethylamine (N-Me-DMPEA)	$\text{C}_{11}\text{H}_{17}\text{NO}_2$	0.69
10	1-Methylbeta-carbolin-7-ol (Harmol)	$\text{C}_{12}\text{H}_{10}\text{N}_2\text{O}$	0.71
11	2-(Methylamino)-1-phenylpropan-1-one (Methcathinone)	$\text{C}_{10}\text{H}_{12}\text{NO}$	0.85
12	1-(p-Fluorophenyl)piperazine (4FPP)	$\text{C}_{10}\text{H}_{13}\text{FN}_2$	0.87
13	2-Phenylethylamine	$\text{C}_8\text{H}_{11}\text{N}$	0.89
14	2-Dimethylamino-1-phenylpropan-1-one (Dimethylcathinone)	$\text{C}_{11}\text{H}_{15}\text{NO}$	0.9
15	1-(2-Methoxyphenyl)piperazine (2MPP)	$\text{C}_{11}\text{H}_{16}\text{N}_2\text{O}$	0.98
16	2,5-Dimethoxyphenethylamine (2C-H)	$\text{C}_{10}\text{H}_{15}\text{NO}_2$	1.00
17	1-(3-Methoxyphenyl)piperazine (3MPP)	$\text{C}_{11}\text{H}_{16}\text{N}_2\text{O}$	1.01
18	2-Ethylamino-1-phenyl-propan-1-one (Ethcathinone)	$\text{C}_{11}\text{H}_{15}\text{NO}$	1.02
19	1-(3-Fluorophenyl)piperazine (3FPP)	$\text{C}_{10}\text{H}_{13}\text{FN}_2$	1.03
20	1-(4-Methoxyphenyl)piperazine (4MPP)	$\text{C}_{11}\text{H}_{16}\text{N}_2\text{O}$	1.04
21	1-(2-Fluorophenyl)piperazine (2FPP)	$\text{C}_{10}\text{H}_{13}\text{FN}_2$	1.07
22	1-(6-Methoxy-2H-benzo[d]1,3-dioxolan-5-yl)prop-2-ylamine (MMDA-2)	$\text{C}_{11}\text{H}_{15}\text{NO}_3$	1.12
23	[2-(5-Methoxyindol-3-yl)ethyl]dimethylamine (5-MeO-DMT)	$\text{C}_{13}\text{H}_{18}\text{N}_2\text{O}$	1.13
24	1-(2-Fluorophenyl)prop-2-ylamine (2FMP)	$\text{C}_9\text{H}_{12}\text{FN}$	1.14
25	2,5-Dimethoxy-4-nitrophenethylamine (2C-N)	$\text{C}_{10}\text{H}_{14}\text{N}_2\text{O}_4$	1.22
26	(2-Indol-3-ylethyl) dimethylamine, dimethyltryptamine (DMT)	$\text{C}_{12}\text{H}_{16}\text{N}_2$	1.23
27	(Indol-3-ylmethyl)dimethylamine (Gramine)	$\text{C}_{11}\text{H}_{14}\text{N}_2$	1.26
28	2-Amino-1-(3,4-methylenedioxy phenyl)butane (BDB)	$\text{C}_{11}\text{H}_{15}\text{NO}_2$	1.29
29	[2-(4-Methoxyphenyl)-isopropyl]methylamine (PMMA)	$\text{C}_{11}\text{H}_{17}\text{NO}$	1.3
30	1-(3-Fluorophenyl)prop-2-ylamine (3FMP)	$\text{C}_9\text{H}_{12}\text{FN}$	1.31
31	4-Fluoro-N-hydroxyamphetamine (N-HO-4FMP)	$\text{C}_9\text{H}_{12}\text{FNO}$	1.32
32	N,N-Diethyltryptamine (DET)	$\text{C}_{14}\text{H}_{20}\text{N}_2$	1.35
33	Methylbeta-carboline (Harman)	$\text{C}_{12}\text{H}_{10}\text{N}_2$	1.36
34	N-Methyl-1-(1,3-benzodioxol-5-yl)-2-butanamine (MBDB)	$\text{C}_{12}\text{H}_{17}\text{NO}_2$	1.44
35	1-(o-Chlorophenyl)piperazine (2CPP)	$\text{C}_{10}\text{H}_{13}\text{ClN}_2$	1.45
36	1-(3-Chlorophenyl)piperazine (3CPP)	$\text{C}_{10}\text{H}_{13}\text{ClN}_2$	1.49
37	3-{2-[Bis(methylethyl)amino]ethyl} indol-4-ol (4-HO-DIPT)	$\text{C}_{16}\text{H}_{24}\text{N}_2\text{O}$	1.53
38	2-Ethylamino-1-(4-methoxyphenyl)propane (PMEA)	$\text{C}_{12}\text{H}_{19}\text{NO}$	1.54
39	1-(2,3-Dimethylphenyl)piperazine (2 3-XP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	1.59
40	Diethyl[2-(5-methoxyindol-3-yl)ethyl]amine (5-MeO-DET)	$\text{C}_{15}\text{H}_{22}\text{N}_2\text{O}$	1.60
41	1-(3-Trifluoromethylphenyl)piperazine (4-TFMPP)	$\text{C}_{11}\text{H}_{13}\text{F}_3\text{N}_2$	1.63
42	[2-(5-Methoxyindol-3-yl)ethyl]methyl (methylethyl)amine (5-MeO-MIPT)	$\text{C}_{15}\text{H}_{22}\text{N}_2\text{O}$	1.64
43	4-Chloro-2,5-dimethoxyphenethylamine (2C-C)	$\text{C}_{10}\text{H}_{14}\text{ClNO}_2$	1.65
44	7-Methoxy-1-methylbeta-carboline (Harmine)	$\text{C}_{13}\text{H}_{12}\text{N}_2\text{O}$	1.67

TABLE 1: Continued.

No.	Name	Structure	k'_{EXP}
45	1-(2,4-Dimethylphenyl)piperazine (2,4-XP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	1.71
46	3-{2-[Methyl (methylethyl)amino]ethyl} indol-4-ol (4HO-MIPT)	$\text{C}_{14}\text{H}_{20}\text{N}_2\text{O}$	1.72
47	4-Bromo-2,5-dimethoxyphenethylamine (2C-B)	$\text{C}_{10}\text{H}_{14}\text{BrNO}_2$	1.73
48	(2-Indol-3-ylethyl)methyl (methylethyl)amine (MIPT)	$\text{C}_{14}\text{H}_{20}\text{N}_2$	1.75
49	1-(4-Chloro-2,5-dimethoxy phenyl)prop-2-ylamine (DOC)	$\text{C}_{11}\text{H}_{16}\text{ClNO}_2$	1.83
50	1-(3-Thienylmethyl)piperazine (3TMP)	$\text{C}_9\text{H}_{14}\text{N}_2\text{S}$	1.89
51	4-Iodo-2,5-dimethoxyphenethylamine (2C-I)	$\text{C}_{10}\text{H}_{14}\text{INO}_2$	1.90
52	3-{2-[Methyl(methylethyl)- amino]ethyl} indo-4-yl acetate (4-AcO-MIPT)	$\text{C}_{16}\text{H}_{22}\text{N}_2\text{O}_2$	1.92
53	Bis(methylethyl)[2-(5-methoxyindol-3-yl)ethyl]amine (5-MeO-DIPT)	$\text{C}_{17}\text{H}_{26}\text{N}_2\text{O}$	1.97
54	3,4-Trimethyleneamphetamine (3,4-TMA)	$\text{C}_{12}\text{H}_{17}\text{N}$	2.00
55	2,5-Dimethoxy-4-ethylphenethylamine (2C-E)	$\text{C}_{12}\text{H}_{19}\text{NO}_2$	2.01
56	1-BenzyM-methylpiperazine (MBZP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	2.03
57	2,5-Dimethoxy-4-ethylthio amphetamine (ALEPH-2)	$\text{C}_{13}\text{H}_{21}\text{NO}_2\text{S}$	2.04
58	1-Benzylpiperazine (BZP)	$\text{C}_{11}\text{H}_{16}\text{N}_2$	2.06
59	2-(4-Iodo-2,5-dimethoxyphenyl)ethylamine (DOI)	$\text{C}_{11}\text{H}_{16}\text{INO}_2$	2.07
60	Bis(methylethyl)(2-indol-3-ylethyl)amine (DIPT)	$\text{C}_{16}\text{H}_{24}\text{N}_2$	2.08
61	1-(2-Phenylethyl)piperazine (2PEP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	2.19
62	2,5-Dimethoxy-4-propylthiophenethylamine (2C-T-7)	$\text{C}_{13}\text{H}_{21}\text{NO}_2\text{S}$	2.22
63	[2-(5-Methoxyindol-3-yl)ethyl]dipropylamine (5-MeO-DPT)	$\text{C}_{17}\text{H}_{26}\text{N}_2\text{O}$	2.45
Prediction set			
64	N,N-Dimethylaminotyramine (Hordenine)	$\text{C}_{10}\text{H}_{15}\text{N}_1\text{O}_1$	0.44
65	2-(4-(2-Aminoethyl)-2,5-dimethoxy phenylthio)ethanol (2C-T-2-OH)	$\text{C}_{12}\text{H}_{19}\text{NO}_3\text{S}$	0.56
66	2-Methylamino-1-(3,4-methylene dioxyphenyl)propan-1-one (Mlone)	$\text{C}_{11}\text{H}_{13}\text{NO}_3$	0.78
67	2,4,5-Trimethoxyamphetamine (TMA-2)	$\text{C}_{12}\text{H}_{19}\text{NO}_3$	0.89
68	2-(5-Methoxyindol-3-yl)ethylamine (5MeO-T)	$\text{C}_{11}\text{H}_{14}\text{N}_2\text{O}$	0.92
69	2-Indol-3-ylethylamine (T)	$\text{C}_{10}\text{H}_{12}\text{N}_2$	1.00
70	N-Methyl-3,4-methylenedioxyamphetamine (MDMA)	$\text{C}_{11}\text{H}_{15}\text{NO}_2$	1.10
71	1-(5-Methoxyindol-3-yl)prop-2-ylamine (5-MeO-AMT)	$\text{C}_{12}\text{H}_{16}\text{N}_2\text{O}$	1.13
72	(2-Indol-3-ylethyl)methylamine (NMT)	$\text{C}_{11}\text{H}_{14}\text{N}_2$	1.17
73	Methyl(1-methyl-2-phenylethyl)amine, Methamphetamine (MA)	$\text{C}_{10}\text{H}_{15}\text{N}$	1.29
74	4-(2H-Benzo[d]1,3-dioxolan-5-yl)- 3-methylbutan-2-ol (FLEA)	$\text{C}_{11}\text{H}_{15}\text{NO}_3$	1.38
75	[2-(4-Fluorophenyl)-isopropyl]methylamine (N-Me-4FMP)	$\text{C}_{10}\text{H}_{14}\text{FN}$	1.48
76	1-(p-Chlorophenyl)piperazine (4CPP)	$\text{C}_{10}\text{H}_{13}\text{ClN}_2$	1.50
77	1-(3,4-Dimethylphenyl)piperazine (3,4-XP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	1.59
78	7-Methoxy-1-methyl- 3,4-dihydrobetacarboline (Harmaline)	$\text{C}_{13}\text{H}_{14}\text{N}_2\text{O}$	1.67
79	1-(5-Trifluoromethylphenyl)piperazine (TFMPP)	$\text{C}_{11}\text{H}_{13}\text{F}_3\text{N}_2$	1.72
80	N-Ethyl-N-(2-(5- methoxy-li/-indol-3- yl)ethyl)propan-2-amine (5-MeO-EIPT)	$\text{C}_{16}\text{H}_{24}\text{N}_2\text{O}$	1.81
81	3-{2-[Bis(methylethyl)-amino]ethyl} indol-4-yl acetate (4-AcO-DIPT)	$\text{C}_{18}\text{H}_{26}\text{N}_2\text{O}_2$	1.95
82	[2-(5-Methoxyindol-3-yl)ethyl]diprop-2-enylamine (5-MeO-DALT)	$\text{C}_{17}\text{H}_{22}\text{N}_2\text{O}$	2.06
83	2,5-Dimethoxy-4-isopropylthio phenethylamine (2C-T-4)	$\text{C}_{13}\text{H}_{21}\text{NO}_2\text{S}$	2.15
84	(2-Indol-3-ylethyl)dipropylamine (DPT)	$\text{C}_{16}\text{H}_{24}\text{N}_2$	2.37
Test set			
85	4-(2-Aminoethyl)phenol (Tyramine)	$\text{C}_8\text{H}_{11}\text{NO}$	0.36
86	p-Hydroxy-methamphetamine (4HO-MA)	$\text{C}_{10}\text{H}_{15}\text{NO}$	0.50
87	(2R)-2-Amino-3-indol-3-ylpropanoic acid (TP)	$\text{C}_{11}\text{H}_{12}\text{N}_2\text{O}_2$	0.60

TABLE 1: Continued.

No.	Name	Structure	k'_{EXP}
88	3,4,5-Trimethoxyamphetamine (TMA)	$\text{C}_{12}\text{H}_{19}\text{NO}_3$	0.87
89	2-Ethylamino-1-(3,4-methylenedioxyphenyl)propan-1-one (bk-MDEA)	$\text{C}_{12}\text{H}_{15}\text{NO}_3$	0.95
90	1-Phenylpiperazine (PP)	$\text{C}_{10}\text{H}_{14}\text{N}_2$	1.00
91	2-Methylamino-1-(3,4-methylenedioxy phenyl)butan-1-one (bk-MBDB)	$\text{C}_{12}\text{H}_{15}\text{NO}_3$	1.04
92	1-Phenylprop-2-ylamine, Amphetamine (AM)	$\text{C}_9\text{H}_{13}\text{N}$	1.13
93	1-Indol-3-ylprop-2-ylamine (AMT)	$\text{C}_{11}\text{H}_{14}\text{N}_2$	1.25
94	1-(4-Fluorophenyl)prop-2-ylamine, 4-Fluoromethamphetamine (4FMP)	$\text{C}_9\text{H}_{12}\text{FN}$	1.34
95	(3-(2H-Benzo[d] 1,3-dioxolan-5-yl)-1- methylpropyl)methylamine (HMDMA)	$\text{C}_{12}\text{H}_{17}\text{NO}_2$	1.45
96	1-(4-Fluorophenyl)-2-methyl-2-propylamine (4F-Phentermine)	$\text{C}_{10}\text{H}_{14}\text{FN}$	1.54
97	2,4,6-Trimethoxyamphetamine (TMA-6)	$\text{C}_{12}\text{H}_{19}\text{NO}_3$	1.61
98	1-(2,5-Dimethylphenyl)piperazine (2,5-XP)	$\text{C}_{12}\text{H}_{18}\text{N}_2$	1.69
99	1-(2H-Benzo[d] 1,3-dioxolan-5-yl)- 2-pyrrolidinylpentan-1-one (MDPV)	$\text{C}_{16}\text{H}_{21}\text{NO}_3$	1.73
100	2,5-Dimethoxy-4-ethylthiophenethylamine (2C-T-2)	$\text{C}_{12}\text{H}_{19}\text{NO}_2\text{S}$	1.87
101	2-[(tert-Butyl)amino]-1-(3-chloro phenyl)propan-1-one (Bupropion)	$\text{C}_{13}\text{H}_{18}\text{ClNO}$	1.98
102	2-Aminopropiophenone (Cathinone)	$\text{C}_9\text{H}_{11}\text{NO}$	2.05
103	Piperonylpiperazine (MDBZP)	$\text{C}_{12}\text{H}_{16}\text{N}_2\text{O}_2$	2.08
104	2-(2,4,5-Trichloro-3,6-dimethoxy phenyl)ethylamine (2C-C-3)	$\text{C}_{10}\text{H}_{12}\text{Cl}_3\text{NO}_2$	2.23

factor (k') ratio of each drug with the internal standard (IS), the ultraviolet (UV) spectrum, and the MS data.

2.2. Computer Hardware and Software. All calculations were run on an HP laptop computer with an AMD Turion64X2 processor and a Windows XP operating system. The optimizations of molecular structures were done by HyperChem 7.0, and descriptors were calculated by Dragon Version 3.0 software. Cross-validation, GA-PLS, L-M ANN, and other calculations were performed in the MATLAB (Version 7, MathWorks, Inc.) environment.

2.3. Molecular Modeling and Theoretical Molecular Descriptors. The derivation of theoretical molecular descriptors proceeds from the chemical structure of the compounds. In order to calculate the theoretical descriptors, molecular structures were constructed with the aid of HyperChem version 7.0. The final geometries were obtained with the semiempirical AM1 method in HyperChem program. The molecular structures were optimized using Fletcher-Reeves algorithm until the root mean square gradient was 0.01 kcal mol⁻¹. The resulted geometry was transferred into Dragon program, to calculate 1497 descriptors, which was developed by Todeschini et al. [21].

2.4. Genetic Algorithm for Descriptor Selection. To select the most relevant descriptors with GA, the evolution of the population was simulated [22–24]. Each individual of the population, defined by a chromosome of binary values, represented a subset of descriptors. The number of the genes at each chromosome was equal to the number of

the descriptors. The population of the first generation was selected randomly. A gene was given the value of one, if its corresponding descriptor was included in the subset; otherwise, it was given the value of zero. The number of the genes with the value of one was kept relatively low to have a small subset of descriptors [25] that is the probability of generating zero for a gene was set greater. The operators used here were crossover and mutation. The application probability of these operators was varied linearly with a generation renewal. For a typical run, the evolution of the generation was stopped, when 90% of the generations had taken the same fitness. In this paper, size of the population is 30 chromosomes, the probability of initial variable selection is 5 : V (V is the number of independent variables), crossover is multipoint, the probability of crossover is 0.5, mutation is multipoint, the probability of mutation is 0.01, and the number of evolution generations is 1000. For each set of data, 3000 runs were performed.

2.5. Nonlinear Model

2.5.1. Artificial Neural Network. An artificial neural network (ANN) with a layered structure is a mathematical system that stimulates biological neural network, consisting of computing units named neurons and connections between neurons named synapses [26–28]. All feed-forward ANN used in this paper are three-layer networks. Each neuron in any layer is fully connected with the neurons of a succeeding layer. Figure 2 shows an example of the architecture of such ANN. The Levenberg-Marquardt backpropagation algorithm was used for ANN training, and the linear functions were used as the transformation functions in hidden and output layers.

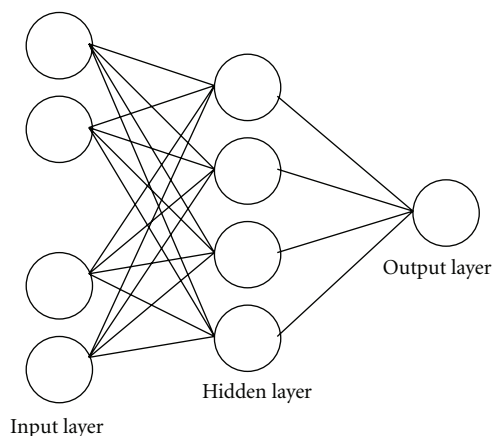


FIGURE 2: Used three-layer ANN.

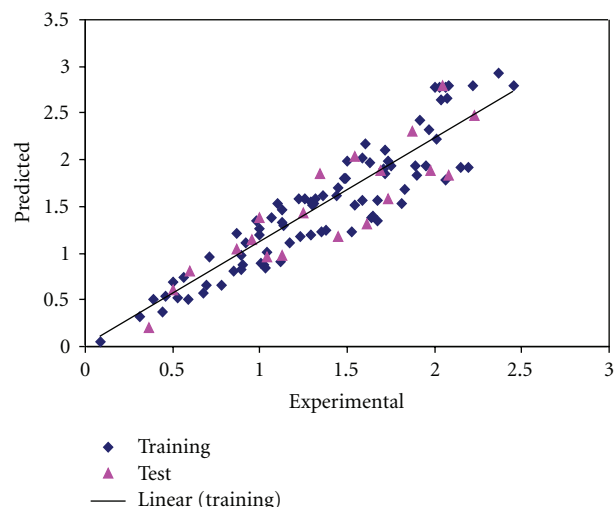
3. Results and Discussion

3.1. Linear Model

3.1.1. Results of the GA-PLS Model. To reduce the original pool of descriptors to an appropriate size, the objective descriptor reduction was performed using various criteria. Reducing the pool of descriptors eliminates those descriptors which contribute either no information or whose information content is redundant with other descriptors present in the pool. These descriptors were employed to generate the models with the GA-PLS program. The best model is selected on the basis of the highest square correlation coefficient leave-group-out cross-validation (R^2), the least root mean squares error (RMSE), and relative error (RE) of prediction. These parameters are probably the most popular measure of how well a model fits the data. The best GA-PLS model contains 11 selected descriptors in 5 latent variables space. The R^2 , mean RE, and RMSE for training and test sets were (0.856, 0.806), (14.11, 20.37), and (0.28, 0.41), respectively. For this in general, the number of components (latent variables) is less than the number of independent variables in PLS analysis. The predicted values of k' are plotted against the experimental values for training and test sets in Figure 3. For this in general, the number of components (latent variables) is less than the number of independent variables in PLS analysis. The PLS model uses higher number of descriptors that allow the model to extract better structural information from descriptors to result in a lower prediction error.

3.2. Nonlinear Model

3.2.1. Results of the L-M ANN Model. The networks were generated using the eleven descriptors appearing in the GA-PLS models as their inputs and k' as their output. For ANN generation, data set was separated into three groups: calibration, prediction, and test sets. A three-layer network with a sigmoid transfer function was designed for each ANN. Before training the networks, the input and output values were normalized between -1 and 1 . The network was

FIGURE 3: Plots of predicted k' against the experimental values by GA-PLS model.

then trained using the training set by the back-propagation strategy for optimization of the weights and bias values. The proper number of nodes in the hidden layer was determined by training the network with different number of nodes in the hidden layer. The root-mean-square error (RMSE) value measures how good the outputs are in comparison with the target values. It should be noted that for evaluating the overfitting, the training of the network for the prediction of k' must stop when the RMSE of the prediction set begins to increase while RMSE of calibration set continues to decrease. Therefore, training of the network was stopped when overtraining began. All of the above-mentioned steps were carried out using basic backpropagation, conjugate gradient and Levenberg Marquardt weight update functions. It was realized that the RMSE for the training and test sets is minimum when seven neurons were selected in the hidden layer. Finally, the number of iterations was optimized with the optimum values for the variables. The R^2 , mean RE, and RMSE for calibration, prediction, and test sets were (0.923, 0.912, 0.880), (8.04, 9.61, 12.08), and (0.16, 0.18, 0.21), respectively. The statistical parameters R^2 , RE, and RMSE were obtained for proposed models. Inspection of the results reveals a higher R^2 and lowers other values parameter for the training, test and prediction sets compared with their counterparts for GA-PLS. Plots of predicted k' versus experimental k' values by L-M ANN for training and test sets are shown Figures 4(a) and 4(b). Obviously, there is a close agreement between the experimental and predicted k' and the data represent a very low scattering around a straight line with respective slope and intercept close to one and zero. This clearly shows the strength of L-M ANN as a nonlinear feature selection method. The key strength of L-M ANN is their ability to allow for flexible mapping of the selected features by manipulating their functional dependence implicitly. The whole of these data clearly displays a significant improvement of the QSRR model consequent to nonlinear statistical treatment.

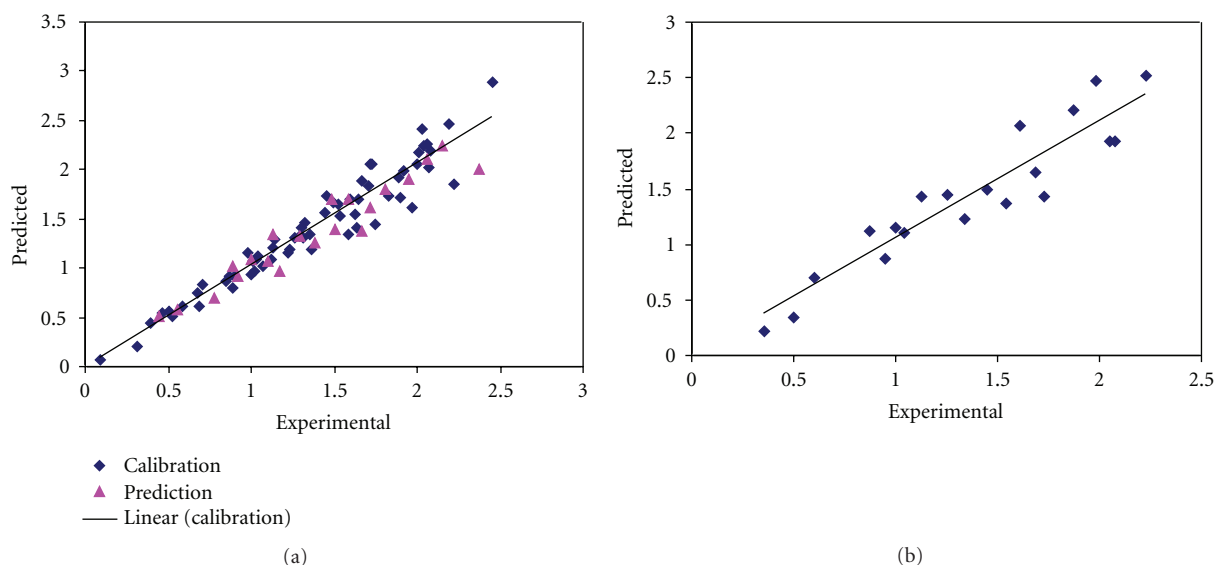


FIGURE 4: Plot of predicted k' obtained by L-M ANN against the experimental values (a) calibration and prediction set of molecules and (b) for test set.

3.3. Model Validation and Statistical Parameters. The applied internal (leave-group-out cross-validation (LGO-CV)) and external (test set) validation methods were used for the predictive power of models. In the leave-group-out procedure one compound was removed from the data set, and the model was trained with the remaining compounds and used to predict the discarded compound. The process was repeated for each compound in the data set. The predictive power of the models developed on the selected training set is estimated on the predicted values of test set chemicals. The data set should be divided into three new subdatasets, one for calibration and prediction (training), and the other one for testing. The calibration set was used for model generation. The prediction set that was applied deal with overfitting of the network, whereas test set whose molecules have no role in model building was used for the evaluation of the predictive ability of the models for external set.

On the other hand, by means of training set, the best model is found, and then, the prediction power of it is checked by test set, as an external data set. In this work, from all 104 components, 63 components are in calibration set, 21 components are in prediction set and 20 components are in test set.

The result clearly displays a significant improvement of the QSRR model consequent to nonlinear statistical treatment and a substantial independence of model prediction from the structure of the test molecule. In the above analysis, the descriptive power of a given model has been measured by its ability to predict partition of unknown designer drugs.

For the constructed models, some general statistical parameters were selected to evaluate the predictive ability of the models for k' values. In this case, the predicted k' of each sample in prediction step was compared with the experimental acidity constant. Root mean square error (RMSE) is a measurement of the average difference between predicted

and experimental values, at the prediction step. RMSE can be interpreted as the average prediction error, expressed in the same units as the original response values. The RMSE was obtained by the following formula:

$$\text{RMSE} = \left[\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]^{1/2}. \quad (2)$$

The second statistical parameter was relative error (RE) that shows the predictive ability of each component and is calculated as

$$\text{RE}(\%) = 100 \times \left[\frac{1}{n} \sum_{i=1}^n \frac{(y_i^{\wedge} - y_i)}{y_i} \right]. \quad (3)$$

The predictive ability was evaluated by the square of the correlation coefficient (R^2) which is based on the prediction error sum of squares and was calculated by the following equation:

$$R^2 = \frac{\sum_{i=1}^n (y_i^{\wedge} - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})}, \quad (4)$$

where y_i is the experimental k' in the sample i , y_i^{\wedge} represented the predicted k' in the sample i , \bar{y} is the mean of experimental k' in the prediction set, and n is the total number of samples used in the test set.

The main aim of the present work was to assess the performances of GA-PLS and L-M ANN for modeling the capacity factor of hazardous psychoactive designer drugs. The procedures of modeling including descriptor generation, splitting of the data, variable selection, and validation were the same as those performed for modeling of the k' of psychoactive designer drugs.

4. Conclusion

The QSRR model provides significant additional insight into the relationship between the molecular structure and fundamental processes and phenomena in chemistry. In this study, the GA-PLS and L-M ANN modeling were applied for the prediction of the capacity factor of hazardous psychoactive designer drugs. Two methods seemed to be useful, although a comparison between methods revealed the slight superiority of the L-M ANN to GA-PLS. High-correlation coefficients training (calibration and prediction) and test sets and low prediction errors confirmed the good predictability of two models for new compounds with very low prediction error (specially for L-M ANN). It is easy to notice that there was a good prospect for the L-M ANN application in the QSRR modeling.

References

- [1] C. Sauer, K. Hoffmann, U. Schimmel, and F. T. Peters, "Acute poisoning involving the pyrrolidinophenone-type designer drug 4'-methyl- α -pyrrolidinohexanophenone (MPPH)," *Forensic Science International*, vol. 208, no. 1–3, pp. e20–e25, 2011.
- [2] I. Vardakou, C. Pistos, and C. Spiliopoulou, "Drugs for youth via internet and the example of mephedrone," *Toxicology Letters*, vol. 201, no. 3, pp. 191–195, 2011.
- [3] T. J. Kauppila, A. Flink, M. Haapala et al., "Desorption atmospheric pressure photoionization-mass spectrometry in routine analysis of confiscated drugs," *Forensic Science International*, vol. 210, no. 1–3, pp. 206–212, 2011.
- [4] J. C. Reepmeyer, J. T. Woodruff, and D. A. d'Avignon, "Structure elucidation of a novel analogue of sildenafil detected as an adulterant in an herbal dietary supplement," *Journal of Pharmaceutical and Biomedical Analysis*, vol. 43, no. 5, pp. 1615–1621, 2007.
- [5] B. H. Chen, J. T. Liu, W. X. Chen et al., "A general approach to the screening and confirmation of tryptamines and phenethylamines by mass spectral fragmentation," *Talanta*, vol. 74, no. 4, pp. 512–517, 2008.
- [6] L. F. Martins, M. Yegles, H. Chung, and R. Wennig, "Sensitive, rapid and validated gas chromatography/negative ion chemical ionization-mass spectrometry assay including derivatization with a novel chiral agent for the enantioselective quantification of amphetamine-type stimulants in hair," *Journal of Chromatography B*, vol. 842, no. 2, pp. 98–105, 2006.
- [7] S. D. Brandt, S. Freeman, I. A. Fleet, P. McGagh, and J. F. Alder, "Analytical chemistry of synthetic routes to psychoactive tryptamines Part II. Characterisation of the speeter and anthony synthetic route to *N,N*-dialkylated tryptamines using GC-EI-ITMS, ESI-TQ-MS-MS and NMR," *Analyst*, vol. 130, no. 3, pp. 330–344, 2005.
- [8] R. Gottardo, F. Bortolotti, G. de Paoli, J. P. Pascali, I. Mikšík, and F. Tagliaro, "Hair analysis for illicit drugs by using capillary zone electrophoresis-electrospray ionization-ion trap mass spectrometry," *Journal of Chromatography A*, vol. 1159, no. 1–2, pp. 185–189, 2007.
- [9] T. Matsumoto, R. Kikura-Hanajiri, H. Kamakura, N. Kawahara, and Y. Goda, "Identification of *N*-methyl-4-(3,4-methylenedioxyphenyl)butan-2-amine, distributed as MBDB," *Journal of Health Science*, vol. 52, no. 6, pp. 805–810, 2006.
- [10] T. K. Spratley, P. A. Hays, L. C. Geer, S. D. Cooper, and T. D. McKibben, "Analytical profiles for five "Designer" tryptamines," *Microgram Journal*, vol. 3, no. 1–2, pp. 54–68, 2005.
- [11] K. Doi, M. Miyazawa, H. Fujii, and T. Kojima, "The analysis of the chemical drugs among structural isomer," *Yakugaku Zasshi*, vol. 126, no. 9, pp. 815–823, 2006.
- [12] G. Klopman and H. Zhu, "Recent methodologies for the estimation of *n*-octanol/water partition coefficients and their use in the prediction of membrane transport properties of drugs," *Mini-Reviews in Medicinal Chemistry*, vol. 5, no. 2, pp. 127–133, 2005.
- [13] K. Bodzioch, A. Durand, R. Kaliszan, T. Bączek, and Y. Vander Heyden, "Advanced QSRR modeling of peptides behavior in RPLC," *Talanta*, vol. 81, no. 4–5, pp. 1711–1718, 2010.
- [14] A. Talevi, M. Goodarzi, E. V. Ortiz et al., "Prediction of drug intestinal absorption by new linear and non-linear QSPR," *European Journal of Medicinal Chemistry*, vol. 46, no. 1, pp. 218–228, 2011.
- [15] H. Noorizadeh and M. Noorizadeh, "QSRR-based estimation of the retention time of opiate and sedative drugs by comprehensive two-dimensional gas chromatography," *Medicinal Chemistry Research*. In press.
- [16] G. Carlucci, A. A. D'Archivio, M. A. Maggi, P. Mazzeo, and F. Ruggieri, "Investigation of retention behaviour of non-steroidal anti-inflammatory drugs in high-performance liquid chromatography by using quantitative structure-retention relationships," *Analytica Chimica Acta*, vol. 601, no. 1, pp. 68–76, 2007.
- [17] O. Devos and L. Duponchel, "Parallel genetic algorithm co-optimization of spectral pre-processing and wavelength selection for PLS regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 107, no. 1, pp. 50–58, 2011.
- [18] V. K. Gupta, H. Khani, B. Ahmadi-Roudi, S. Mirakhorli, E. Fereyduni, and S. Agarwal, "Prediction of capillary gas chromatographic retention times of fatty acid methyl esters in human blood using MLR, PLS and back-propagation artificial neural networks," *Talanta*, vol. 83, no. 3, pp. 1014–1022, 2011.
- [19] M. Ferrand, B. Huquet, S. Barbey et al., "Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: interest of applying a variable selection by genetic algorithms before a PLS regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 106, no. 2, pp. 183–189, 2011.
- [20] M. Takahashi, M. Nagashima, J. Suzuki, T. Seto, I. Yasuda, and T. Yoshida, "Creation and application of psychoactive designer drugs data library using liquid chromatography with photodiode array spectrophotometry detector and gas chromatography-mass spectrometry," *Talanta*, vol. 77, no. 4, pp. 1245–1272, 2009.
- [21] R. Todeschini, V. Consonni, A. Mauri et al., "DRAGON-Software for the calculation of molecular descriptors," Version 3.0 for Windows, 2003.
- [22] R. Leardi, "Genetic algorithms," in *Comprehensive Chemometrics*, S. D. Brown, R. Tauler, and B. Walczak, Eds., chapter 1.20, pp. 631–653, Elsevier, 2009.
- [23] S. Riahi, E. Pourbasheer, M. R. Ganjali, and P. Norouzi, "Investigation of different linear and nonlinear chemometric methods for modeling of retention index of essential oil components: concerns to support vector machine," *Journal of Hazardous Materials*, vol. 166, no. 2–3, pp. 853–859, 2009.
- [24] X. Zhou, Z. Li, Z. Dai et al., "QSAR modeling of peptide biological activity by coupling support vector machine with particle swarm optimization algorithm and genetic algorithm," *Journal of Molecular Graphics and Modelling*, vol. 29, no. 2, pp. 188–196, 2010.

- [25] N. Savory, K. Abe, K. Sode, and K. Ikebukuro, "Selection of DNA aptamer against prostate specific antigen using a genetic algorithm and application to sensing," *Biosensors and Bioelectronics*, vol. 26, no. 4, pp. 1386–1391, 2010.
- [26] D. L. Tong and A. C. Schierz, "Hybrid genetic algorithm-neural network: feature extraction for unprocessed microarray data," *Artificial Intelligence in Medicine*, vol. 53, no. 1, pp. 47–56, 2011.
- [27] M. Serati, S. Salvatore, G. Sisto et al., "Urinary symptoms and urodynamic findings in women with pelvic organ prolapse: is there a correlation? results of an artificial neural network analysis," *European Urology*, vol. 60, no. 2, pp. 253–260, 2011.
- [28] R. E. Hoffman, U. Grasmann, R. Gueorgieva, D. Quinlan, D. Lane, and R. Mikkilainen, "Using computational patients to evaluate illness mechanisms in schizophrenia," *Biological Psychiatry*, vol. 69, no. 10, pp. 997–1005, 2011.

