

IDENTIFICAREA AUTOMATĂ A AMFETAMINELOR CU REȚELE NEURONALE ARTIFICIALE - STUDIU COMPARATIV

Asist. drd. fiz. Steluța GOSAV
Universitatea „Dunărea de Jos”, Galați



A absolvit Facultatea de Fizică, Universitatea « Al. I. Cuza », Iași în anul 1993. Este doctorand la Universitatea « Al. I. Cuza » din Iași. A publicat 4 articole ISI, o carte de specialitate, 10 articole publicate în reviste recunoscute de CNCSIS și 20 lucrări comunicate la conferințe naționale și internaționale. Domeniul de cercetare este « Inteligența artificială aplicată în domeniul farmaceutic ».



Prof. dr. fiz. Mirela PRAISLER
Universitatea „Dunărea de Jos”, Galați

A absolvit Facultatea de Fizică, Universitatea București, în anul 1984. Este doctor în Fizică, domeniul Optică și spectroscopie. A publicat 14 articole ISI, 9 cărți de specialitate și are 41 articole prezentate la conferințe naționale și internaționale. A condus 4 proiecte de cercetare științifică internaționale și 6 proiecte naționale. Este specialist în Inteligența artificială aplicată în domeniul farmaceutic.

REZUMAT

În acest articol prezentăm o serie de rețele neuronale artificiale specializate în identificarea amfetaminelor ilicite. Baza de date este formată din substanțe prohibite precum droguri (mai ales stimulante ale SNC, halucinogene, amine simpatomimetice, narcotice și alte posibile analgezice), precursori sau derivați ai acestora. S-au construit trei rețele neuronale artificiale având ca date de intrare spectrele GC-FTIR, descriptorii constituționali și descriptorii topologici ai compușilor din baza de date. Se prezintă o analiză comparativă a rezultatelor validării celor trei rețele neuronale construite. De asemenea se analizează puterea de discriminare/modelare a fiecărui tip de variabilă de intrare folosită.

ABSTRACT

In this article we are presenting three artificial neural networks (ANN) specialized in the identification of illicit amphetamines. The database is composed by prohibited substances such as drugs of abuse (mainly central stimulants, hallucinogens, sympathomimetic amines, narcotics and other potent analgesics), precursors and derivatized counterparts. We have built three neural networks, using as input data the GC-FTIR spectra, constitutional descriptors and topological descriptors computed for all the compounds forming the database. A comparative analysis of the validation results is presented. We are analyzing the power of discrimination/modelling of each ANN system.

Cuvinte cheie: rețea neuronală, amfetamină, descriptor

Keywords: neural network, amphetamine, descriptor

1. INTRODUCERE

Clasa de compuși cu numărul cel mai mare de molecule individuale pe piața ilicită de droguri este clasa feniletilaminelor. Folosirea lor abuzivă este determinată de efectele halucinogene și/sau stimulante [1] pe care acestea le au. Această clasă de compuși este formată dintr-o serie de analogi având drept structură de bază feniletilamina. În scopul de a ieși de sub incidența legii, laboratoarele clandestine modifică structura moleculară a acestor compuși prin adăugarea sau schimbarea unor substituenți din dife-

rite poziții ale structurii de bază, fără a reduce semnificativ efectul psihotrop. Astfel, apariția de noi compuși analogi pune în dificultate specialiștii în toxicologie, care trebuie să identifice noile structuri chimice.

Amfetaminele au o structură chimică de bază care conține un ciclu aromatic legat la o grupare amino prin intermediul unei catene laterale alifatică cu unul (1-feniletilamina) sau doi (2-feniletilamina) atomi de carbon (fig. 1).

Un factor care influențează activitatea toxicologică a analogilor amfetaminei este modul de substituție a ciclului aromatic. Astfel, analogii amfetaminei care conțin grupa-

rea fenil nesubstituită formează clasa de amfetamine care stimulează sistemul nervos central. Analogii amfetaminei care conțin inelul fenilic dublu substituit, de exemplu 3,4-metilendioximetamfetamina (cunoscută pe piața neagră sub numele de „Ecstasy”), produc efecte neurotoxice și halucinații.

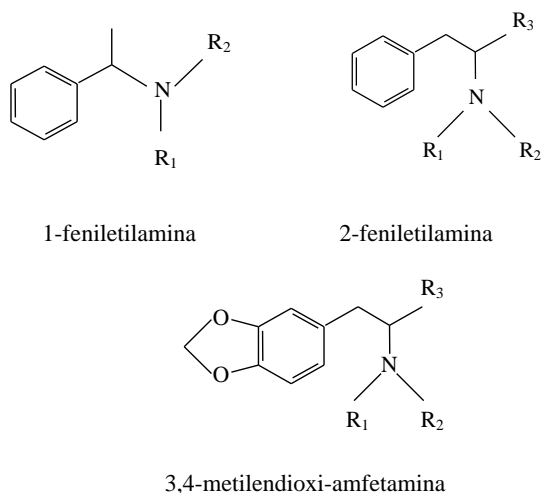


Fig. 1. Structurile moleculare ale principalilor analogi ai amfetaminei.

Spectroscopia în infraroșu (IR) a fost remarcată ca tehnica analitică cu cea mai mare selectivitate datorită posibilității de a determina regiunea „amprentă” a moleculelor. Deoarece amfetaminele sunt substanțe volatile, tehnica GC – FTIR (cromatografie gazoasă cuplată cu spectrometria IR cu transformata Fourier) a devenit una dintre tehnicile cele mai folosite pentru identificarea acestora. În general, spectrele IR sunt foarte complexe, interpretarea lor presupunând multă experiență din partea specialistului. Din acest motiv, interpretarea și identificarea lor automată prin folosirea metodelor numerice (recunoașterea formelor, rețele neuronale artificiale etc.) [2 - 4] a devenit o alternativă din ce în ce mai folosită.

În aceasta lucrare ne propunem să analizăm performanțele câtorva rețele neuronale specializate în identificarea amfetaminelor în cazul folosirii a diferite variabile de intrare: spectre GC-FTIR, descriptori constituționali și descriptori topologici. Totodată, prezentăm o analiză comparativă a rezultatelor validării celor trei rețele neuronale construite. În plus, analizăm puterea de discriminare/modelare a fiecărui tip de variabilă de intrare folosită.

2. PARTEA EXPERIMENTALĂ

2.1. Obținerea bazei de date

Instalația experimentală este formată dintr-un auto-sistem GC Perkin Elmer (Buckinghamshire, UK) dotat cu o interfață GC-IR 2000 și conectat la un sistem FTIR

2000 cu sursa în infraroșu mediu și bandă medie și un detector MCT (telurat de cadmiu și mercur) răcit cu azot lichid. Separarea la temperaturi programate s-a realizat cu o coloană capilară (Hewlett-Packard 5Palo, CA, USA) Ultra-1 din metilsilicon (25 m × 0,32 μm i.d., 0,52 μm grosimea filmului). Gazul purtător a fost heliu la o viteză a pompei de 1,8 ml min⁻¹. Ieșirea coloanei analitice a fost cuplată cu intrarea coloanei iradiate. S-a adăugat heliu ca gaz purtător pentru a menține constant debitul de 1,8 ml min⁻¹ la conexiunea dintre coloana capilară și coloana iradiată. Coloana iradiată (12 cm × 1 mm i.d.) are pe pereții interiori depus un strat subțire de aur și fost încălzită la o temperatură constantă de 270°C. Spectrele înregistrate în timp real au fost obținute prin adăugarea a două scanări cu o rezoluție spectrală de 8 cm⁻¹ și 32 de scanări ale fondului de absorbție. Cromatogramele au fost calculate prin metoda Gram-Schmidt a ortogonalizării vectorilor. Soluțiile metanolice stoc (1,0 mg/ml) ale standardelor de referință au fost injectate în sistemul GC-FTIR. Valoarea concentrației soluțiilor stoc a fost aleasă de 1,0 mg ml⁻¹ deoarece extractele metanolice ale probelor judiciare sunt în același domeniu de concentrație. Reconstrucția Gram-Schmidt a fost realizată utilizând 10 vectori de bază la o scanare. Corecția liniei de bază a fost realizată asupra gaz-cromatogramei reconstruite Gram-Schmidt și spectrele FTIR în faza de vapori (cu zgomot redus) au fost generate după co-adăugare. Spectrele de referință obținute au fost stocate după normalizare într-o bază de date (biblioteca spectrală digitală). Procedura de normalizare presupune scalarea fiecărui spectru astfel încât absorbția celei mai intense benzi să devină egală cu 1. Domeniul de scanare a fost de la 4000 la 580 cm⁻¹. Toate spectrele din bibliotecă au fost reduse prin eliminarea ferestrelor spectrale în care compusul din baza de date nu are absorbții IR. Ca urmare, domeniul IR a fost restrâns de la 3745 la 2555 cm⁻¹ și de la 1995 la 605 cm⁻¹, domeniu care generează 260 de absorbții luate din 10 cm⁻¹ în 10 cm⁻¹.

2.2. Construirea rețelelor neuronale artificiale NN și NNI

Baza de date este formată din 158 de spectre FTIR ale unor compuși reprezentând substanțe prohibite precum droguri (mai ales stimulante ale SNC, halucinogene, amine simpatomimetice, narcotice și alte posibile analgezice), precursori sau derivați ai acestora. Probele constituie standarde de referință precum și compuși sintetizați în laborator. Baza de date este compusă din 15 amfetamine stimulante, 8 amfetamine halucinogene și 135 de compuși diverși incluși în categoria „neamfetamine”.

Rețelele neuronale specializate în identificarea amfetaminelor au fost construite cu ajutorul programului *Easy NNplus*. S-au construit două rețele neuronale cu retropropagare care au număr diferit de variabile de intrare: prima

rețea (rețeaua NN) folosește drept variabile de intrare cele 260 de absorbții, iar a doua rețea (rețeaua NNi) are doar 100 variabile de intrare reprezentând cele mai importante 100 absorbții din numărul total de 260. Ambele rețele sunt structurate pe trei niveluri: un nivel de intrare, unul ascuns și unul de ieșire. Importanța unei variabile de intrare (neuron de pe nivelul de intrare) este dată de suma ponderilor conexiunilor dintre acel neuron și fiecare neuron de pe nivelul următor (nivelul ascuns). Nivelul de ieșire conține trei neuroni, egal cu numărul de clase: clasa amfetaminelor stimulante (clasa M), clasa amfetaminelor halucinogene (clasa T) și clasa neamfetaminelor (clasa N). Pentru fiecare rețea numărul de neuroni de pe nivelul ascuns și ponderile conexiunilor sunt optimizate. În urma procesului de optimizare, pentru rețeaua NN s-a obținut o arhitectură constând dintr-un număr de 12 neuroni ascunși și, respectiv, 3480 conexiuni ponderate, iar pentru rețeaua NNi a rezultat un număr de 13 neuroni ascunși și 1339 conexiuni (fig. 2). Numărul de conexiuni ponderate pentru rețeaua NN fiind mare în comparație cu rețeaua NNi, instruirea și validarea durează foarte mult în primul caz. De aceea a fost reținută pentru analiză doar rețeaua NNi. Instruirea se face cu ajutorul algoritmului de retro-propagare și funcția de transfer folosită este funcția sigmoidă. Valorile ratei de instruire și ale parametrului momentum obținute în urma optimizării sunt: $\alpha = 0,7$ și $\mu = 0,80$.

1	Nivelul	Nivelul
2	ascuns	de ieșire
3		
4		
5		
6		
7		

Pentru ca rețeaua să aibă o capacitate de generalizare cât mai bună în urma procesului de instruire, trebuie ca setul de instruire să conțină cât mai multe exemple reprezentative claselor de atribuire (clasele M, T și N).

În cazul nostru, acest set este alcătuit din 29 de compuși: 7 amfetamine stimulante, 5 amfetamine halucinogene și 17 neamfetamine. Restul de 129 de compuși din baza de date sunt incluși în setul de validare. Stimulantele și halucinogenele din setul de instruire au fost selectate în ordinea similarității lor cu compusul părinte (amfetamina pentru stimulante și 3,4-metilendioxi-amfetamina pentru halucinogene). Ordinea a fost stabilită prin calcularea distanței euclidiene dintre spectrul compusului părinte și spectrul fiecărui analog. Numele stimulantele și halucinogenelor din setul de instruire este: *N*-etilamfetamină, amfetamină, metamfetamină, *N*-*n*-propilamfetamină, β -feniletilamină, α -feniletilamină și *N*-metil- α -feniletilamină, și

3,4-metilendioxi-amfet-amină, 3,4-metilendioxi-metamfetamină, 3,4-metilendioxi-*N*-etilamfetamină, *N*-metil-1-(3,4-metilendioxi-fenil)-2-butanamină și 1-(3,4-metilendioxi-fenil)-2-butanamină. Cele 17 neamfetamine din setul de instruire au fost selectate la întâmplare din baza de date și sunt: prolantan, nicotina, dextromoramida, putrescină, piracetam, nicotamidă, γ – valerolactonă, trimetilsilil (TMS) de γ – acid hidroxicvaleric, trimetilsilil de γ – acid hidroxicbutiric, γ – butir-lactonă, cafeină, codeină și derivatul ei PFP, cadaverină și derivatul ei HFB, β – butir-lactonă și bemegridă.

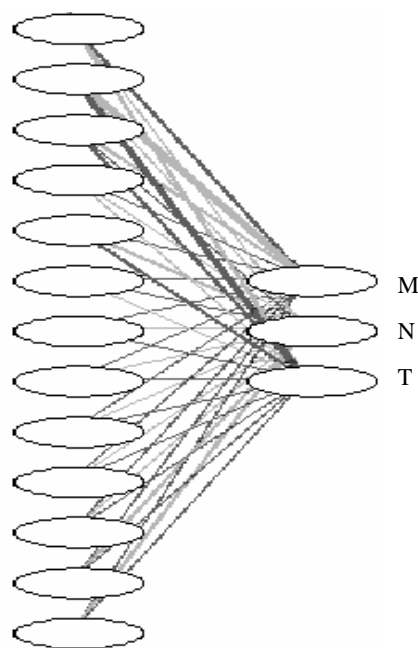


Fig. 2. Nivelele ascunse și de ieșire pentru rețeaua NNi.

2.3. Construirea rețelelor neuronale artificiale CD-NN și TD-NN

Descriptorii moleculari reprezintă elemente importante ce caracterizează structura moleculară a unui compus chimic. Ei conțin informație utilă în descrierea și clasificarea structurilor moleculare legată de proprietățile chimice și biologice ale compușilor respectivi [5, 6].

Baza de date este formată din 146 compuși reprezentând substanțe prohibite precum droguri, precursori sau derivați ai acestora. Moleculele acestor compuși au fost reprezentate în coordonate 3D cu ajutorul programului *HyperChem* (version 6.03). Optimizarea geometriei structurilor moleculare s-a realizat utilizând metoda semi-cuantică PM3. Fișierele care conțin moleculele optimizate constituie intrări pentru programul *Dragon* (version 1.11) care calculează 40 de descriptori constituționali și 69 de descriptori topologici. Descrierea detaliată a acestor descriptori este realizată în Todeschini *et al.* [7].

Din cei 146 de compuși din baza de date sunt: 14 amfetamine stimulante, 7 amfetamine halucinogene și 125 neamfetamine. Rețelele neuronale descrise în acest paragraf au aceeași arhitectură cu rețelele descrise mai sus. În plus, se utilizează același algoritm de învățare și funcție de transfer. Halucinogenele din setul de învățare sunt enumerate mai sus (în paragraful 2.2).

Am construit o rețea neuronală având ca variabile de intrare 40 descriptori constituționali (rețeaua 40CD-NN). Acești descriptori sunt prezentați într-o lucrare anterioară [8]. Setul de învățare este format din 29 de compuși: 6 stimulante, 5 halucinogene și 18 neamfetamine. Stimulantele din setul de învățare sunt: fenproporex, metamfetamina, β -feniletilamina, α -feniletilamina, benzfetamina și *N*-metil- α -feniletilamina. Cele 18 neamfetamine (prolintan, nicotina, ketamina, 4-iod-2,5-dimetoxiamfetamina, putrescina, piracetam, nicotamida, γ – valerolactona, trimetilsilil (TMS) de γ – acid hidroxivaleric, trimetilsilil de γ – acid hidroxibutiric, γ – butirolactona, cafeina, codeina și derivatul ei PFP, cadaverina și derivatul ei HFB, β – butirolactona și bemegrida) au fost alese în urma unui proces de optimizare de tip încercare și eroare [8]. În urma procesului de optimizare rețeaua 40CD-NN are 559 conexiuni ponderate, rata de învățare $\alpha = 0,6$ și parametrul momentum $\mu = 0,70$.

Aceeași bază de date de 146 de compuși a fost utilizată pentru a construi o rețea neuronală ce are ca variabile de intrare descriptori topologici. Din cei 69 de descriptori topologici am ales pe primii 30 cei mai *senzitivi* descriptori. Senzitivitatea unei variabile de intrare este dată de măsura în care o ieșire se schimbă atunci când intrarea variază ca valoare. Cu descriptorii selectați am construit rețeaua 30TDs-NN [9] folosind un set de învățare compus din 28 de probe: 6 stimulante, 5 halucinogene și 17 neamfetamine. Stimulantele din setul de învățare sunt: amfetamina, metamfetamina, β -feniletilamina, α -feniletilamina, benzfetamina și *N*-metil- α -feniletilamina. În urma procesului de optimizare rețeaua 30TDs-NN are 14 neuroni pe nivelul ascuns, 462 conexiuni ponderate, rata de învățare $\alpha = 0,70$ și parametrul momentum $\mu = 0,80$. Neamfetaminele din setul de învățare sunt enumerate în paragraful 2.2.

3. REZULTATE ȘI DISCUȚII

Toate rețelele neuronale au fost programate să oprească procesul de instruire când eroarea medie scade sub valoarea erorii țintă. Aceasta a fost stabilită la valoarea de 0,01.

Pentru a putea realiza o comparație a performanțelor rețelelor neuronale, s-a efectuat validarea tuturor compușilor din baza de date. Metoda de validare folosită este cea încrucișată (*cross-validation*) deoarece numărul de probe din baza de date este relativ mic.

O evaluare a performanțelor celor două rețele se poate face calculând următorii parametri: rata de clasificare (C),

rata de clasificare corectă (CC), rata de pozitive adevărate (TP), rata de negative adevărate (TN), rata de pozitive false (FP) și rata de negative false (FN). Probele « pozitive » sunt amfetaminele, iar probele « negative » sunt neamfetaminele. Rata de clasificare reprezintă procentul de probe din baza de date care sunt clasificate (corect sau incorect). Rata de pozitive adevărate reprezintă procentul de probe pozitive care sunt clasificate corect din numărul total de pozitive (stimulante și halucinogene), iar rata de pozitive false reprezintă procentul de probe pozitive din numărul total de pozitive (stimulante și halucinogene) care sunt clasificate incorect. Tabelul 1 prezintă rezultatele validării.

Clasificarea corectă a probelor pozitive este foarte importantă în practica judiciară, unde identificarea drogurilor nu trebuie să dea greș. O prima remarcă foarte încurajatoare este că toate rețelele neuronale identifică corect amfetaminele, adică rata TP este de 100%.

Tabelul 1. Rezultatele procesului de validare

	NNi	30TDs-NN	40CD-NN
TP(%)	100	100	100
TN(%)	85.07	73.73	84.82
FP(%)	14.93	26.27	15.18
FN(%)	0.0	0.0	0.0
C(%)	98.74	94.52	91.09
CC(%)	87.26	77.54	87.22

Observăm că, indiferent de variabilele de intrare utilizate, are loc o modelare foarte bună a claselor M și T. Rata de clasificare corectă a negativelor pentru rețeaua 30TDs-NN are valoarea cea mai mică (73,73%), pe când pentru rețelele NNi și 40CD-NN, această rată are valori aproximativ apropiate (85,07%, respectiv 84,82%). Aceeași situație apare în cazul ratei de clasificare corecte. Singurul parametru care are valoare mică pentru rețeaua 40CD-NN în comparație cu celelalte două rețele este rata de clasificare, care este 91,09% în cazul 40CD-NN, în timp ce pentru rețeaua 30TDs-NN este 94,52%, iar pentru rețeaua NNi ia valoarea cea mai mare ($C = 98,74\%$). Acest lucru se datorează faptului că rețeaua 40CD-NN beneficiază de mai multă informație de intrare. Însă rata superioară de clasificare a acestei rețele nu compensează dezavantajul unor rate mai nefavorabile de clasificare a pozitivelor și a negativelor adevărate.

4. CONCLUZII

Rețelele neuronale care utilizează absorbții FTIR drept date de intrare dau rezultatele cele mai bune pentru identificarea automată a amfetaminelor în funcție de activitatea lor biologică. Trebuie însă să remarcăm că și descriptorii constituționali dau rezultate bune, respectiv

rate de clasificare / identificare acceptabile pentru practica analizelor toxicologice. Un alt avantaj al rețelei 40CD-NN este că poate fi folosită pentru a testa noi structuri moleculare ce pot fi utilizate în scopuri clinice *înainte* ca acești compuși să fie sintetizați, reducând astfel cheltuielile pentru obținerea de noi amfetamine pentru scopuri medicale legitime. Descriptorii topologici au cea mai mică putere de modelare/discriminare în comparație cu spectrele FTIR și descriptorii constituționali, fiind deci mai puțin recomandați pentru scopul identificării noilor structuri moleculare care ar putea avea un efect farmaceutic similar amfetaminelor biologic active.

BIBLIOGRAFIE

1. **S.B. Karch**, 1998, *Drug Abuse Handbook*, CRC Press, New York,.
2. **M. Praisler, I. Dirinck, J. Van Bocxlaer, A.P. De Leenheer, D.L. Massart**, 2000, Pattern recognition techniques screening for drugs of abuse with gas chromatography – Fourier transform infrared spectroscopy, *Talanta*, Vol 53, p. 177-193.
3. **H. Yang, P.R. Griffiths, J. D. Tate**, 2003, Comparison of partial least squares regression and multi-layer neural networks for quantification of nonlinear systems and application to gas phase Fourier transform infrared spectra, *Anal. Chim. Acta*, Vol 489, p. 125-136.
4. **N. Benoudjit, E. Cools, M. Meurens, M. Verleysen**, 2004, Chemometric calibration of infrared spectrometers: selection and validation of variables by non-linear models, *Chem. and Intell. Lab. Syst.*, Vol 70, p. 47-53.
5. **Q. Li, X. Chen, Z. Hu**, 2004, Quantitative structure–property relationship studies for estimating boiling points of alcohols using calculated molecular descriptors with radial basis function neural networks, *Chem. and Intell. Lab. Syst.*, Vol 72, p. 93-100.
6. **I. Valkova, M. Vracko, S. C. Basak**, 2004, Modeling of structure–mutagenicity relationships: counter propagation neural network approach using calculated structural descriptors, *Anal. Chim. Acta* Vol 509, p. 179-186.
7. **R. Todeschini, V. Consonni**, 2000, *Handbook of molecular descriptors*, Wiley-VCH, Weinheim.
8. **S. Gosav, M. Praisler, D.O. Dorohoi, G. Popa**, 2006, Structure-activity correlations for illicit amphetamines using ANN and constituțional descriptors, *Talanta* Vol 70, p. 922-928.
9. **S. Gosav, M. Praisler, D.O. Dorohoi**, 2007, ANN expert system screening for illicit amphetamines using molecular descriptors, *Journal of Molecular Structure* p. 834-836.