Quantitative Structure–Retention Relationship Study of Benzodiazepines Using Adaptive Neuro Fuzzy Inference System as Feature Selection Method

ARTICLE *in* QSAR & COMBINATORIAL SCIENCE · APRIL 2008

Impact Factor: 1.55 · DOI: 10.1002/qsar.200630163

4 AUTHORS, INCLUDING:

Mehdi Jalali-Heravi
Sharif University of Technology
63 PUBLICATIONS **1,017** CITATIONS
SEE PROFILE

Anahita Kyani
University of Michigan
15 PUBLICATIONS **183** CITATIONS
SEE PROFILE

Sepideh Afsari
Temple University
4 PUBLICATIONS **10** CITATIONS
SEE PROFILE

QSAR & Combinatorial Science QCS

Quantitative Structure – Retention Relationship Study of Benzodiazepines Using Adaptive Neuro Fuzzy Inference System as Feature Selection Method

Mehdi Jalali-Heravi*, Anahita Kyani, Sepideh Afsari-Mamaghani and Atieh Ghadiri-Bidhendi

Department of Chemistry, Sharif University of Technology, P. O. Box 11155 – 9516, Tehran, Iran, E-mail: jalali@sharif.edu

Keywords: Adaptive neuro fuzzy inference system, Artificial neural networks, Benzodiazepines, Multiple linear regression, Quantitative structure – retention relationship

Received: December 6 2006; Accepted: June 19 2007

DOI: 10.1002/qsar.200630163

Abstract

A Quantitative Structure – Retention Relationship (QSRR) study of 32 benzodiazepines is performed in this work. Two feature selection methods of Adaptive Neuro Fuzzy Inference System (ANFIS) and a stepwise regression approach adopted for the Multiple Linear Regressions (MLR) were used to predict the Liquid Chromatography-Mass Spectrometry (LC-MS) Retention Time (RT) of these compounds on a Xterra MS C-18 stationary phase. ANFIS and MLR methods were used as variable selection tools and a neural network was employed for predicting the RTs. Three descriptors of 3D-MoRSE-signal 06/weighted by atomic polarizabilities (Mor06p), Radial Distribution Function-1.0/weighted by atomic van der Waals volumes (RDF010v), and number of functional groups of R−Co−N< and >N−C=N (N-072) reveal the importance of dispersion interactions between benzodiazepines and C-18 stationary phases and also electrostatic and hydrogen bond interactions of these compounds with the polar mobile phase. The superiority of 3-4-1 ANFIS-Artificial Neural Networks (ANN) model over MLR-ANN with a linear feature selection method was illustrated by Leave-Multiple-Out (LMO) cross-validation method. Values of $R^2_{L6O} = 0.771$ and $RMSE_{L6O} = 4.792$ for ANFIS-ANN model should be compared with the values of $R^2_{L6O} = 0.632$ and $RMSE_{L6O} = 6.481$ for the MLR-ANN model. LMO-CV and Y-randomization results indicate the robustness and predictive ability of the generated model of ANFIS-ANN and show the ability of ANFIS in selecting the features for phenomena with nonlinear characteristics. Sequential Zeroing of Weights (SZW) as a sensitivity analysis method showed the importance of dispersion interaction in the retention mechanism of benzodiazepines in LC method.

1 Introduction

Benzodiazepines are substances with a broad range of therapeutic uses. Their medical use varies, but they are predominantly used as hypnotics and sedatives. Some members are also used in the treatment of post-traumatic stress and obsessive – compulsive disorders, alcohol withdrawal, muscle spasm, and seizures [1]. They are also commonly used as "date-rape" drugs to render a victim incapable of resisting an attack. The benzodiazepines are relatively safe drugs with mild side effects. However, elderly patients are at an increased risk of respiratory depression of benzodiazepine [1, 2].

Benzodiazepines elicit a large number of physiological and psychological responses in humans that can often lead to significant behavioral changes and adverse effects on

Abbreviations: **ANFIS**, adaptive neuro fuzzy inference system; **ANN**, artificial neural networks; **BP-ANN**, back propagation artificial neural networks; **E2m**, 2nd component accessibility directional WHIM index/weighted by atomic masses; **FL**, fuzzy logic; **FIS**, fuzzy inference system; **LC-MS**, liquid chromatography-mass spectrometry; **LMO-CV**, leave-multiple-out cross-validation; **LOO-CV**, leave-one-out cross-validation; **L3O**, leave-3-out; **L6O**, leave-6-out; **MFs**, membership functions; **MLR**, multiple linear regression; **Mor06p**, 3D-MoRSE-signal 06/weighted by atomic polarizabilities; **NNs**, neural networks; **N-072**, number of functional groups of R−Co−N< and >N−C=N; **RDF010v**, radial distribution function- 1.0/weighted by atomic van der Waals volumes; **R4e+**, R maximal autocorrelation of lag 4/weighted by atomic Sanderson electronegativities; **RMSE**, root mean squared error; $\mathbf{RMSE_{CV}}$, root mean squared error of cross-validation; **RT**, retention time; **SEP**, standard error of prediction; **SET**, standard error of training; **SZW**, sequential zeroing of weights

skills required for safe driving [3]. These include reduced lane control, increased reaction times, reduced hand-eye coordination, and cognitive impairment. Impairment can exceed more than what has been seen with 0.05 g% ethanol. In high doses, benzodiazepines can cause persons to exhibit classical features of CNS-depressant drugs such as nystagmus, ataxia, slurred speech, and impaired divided attention skills [1].

During recent years, substantial research has been focused on the development of identification and quantification procedures for drugs in human blood using Liquid Chromatography-Mass Spectrometry (LC-MS) [4, 5]. Analysis of these drugs in blood samples may be indicated in a lot of forensic cases such as driving under the influence of drugs, cases of date-rape or violent crime, and cases of unknown causes of death.

Despite many experimental works published on developing analytical techniques for screening and quantifying benzodiazepines [6, 7], the literature does not contain papers on the use of computer-assisted methods for the modeling of chromatographic parameters of such an important class of drugs.

As a consequence, development of an accurate and versatile theoretical model for predicting the LC-MS Retention Time (RT) of benzodiazepines seems to be very useful. Recently, using hybrid chemometric methods is common for predicting different properties and selecting the most important parameters affecting the property of interest [8, 9]. The present paper reports on the usefulness of Adaptive Neuro Fuzzy Inference System (ANFIS) as a feature selection tool in combination with the Artificial Neural Network (ANN) method in predicting and modeling of the RT of benzodiazepines, their active metabolites, and benzodiazepine-like substances obtained by LC-MS spectrometry [2]. The main goals of the present work were: (1) to accurately predict the RT of benzodiazepines on a nonpolar stationary phase of C-18 and a polar mobile phase of methanol/formic acid, (2) to achieve a better understanding of the physico – chemical basis of retention behavior of benzodiazepines and its mechanism, (3) to consider the linear or nonlinear characteristic of the retention of these compounds on nonpolar C-18 stationary phase, (4) to compare the ability of two feature selection methods of stepwise regression approach adopted for Multiple Linear Regression (MLR) and ANFIS in selecting the inputs for a nonlinear chemometric technique of neural networks, and (5) to study the ANFIS selected variable influence and contribution to the retentions using Sequential Zeroing of Weights (SZW) as a variable sensitivity analysis method.

The results of the present work show that the ANFIS-ANN model can predict the LC-MS RT of benzodiazepines by using three interpretable descriptors and reduce the need for different time-consuming, difficult, and expensive stages of experiments.

2 Methods

The main aim of the present work was development of an artificial neural net to predict the RT of benzodiazepines. One of the main challenges in developing these types of models is choosing adequate descriptors as their inputs. There are two different methods of feature selection technique: objective and subjective methods. The former method selects the descriptors based on the relationship between them, whereas the latter one uses the relation between the descriptors and the dependent variable, *i.e.*, RT for the variable selection.

The stepwise regressions adopted for MLR and ANFIS were employed as subjective feature selection techniques. The ANFIS has been used as a nonlinear tool to select descriptors as inputs of ANN. Stepwise regression adopted for MLR was used as the linear subjective one. Therefore, the descriptors that appeared in the ANFIS and MLR model were used as inputs for developing the ANN. The best generated MLR model was also considered as the calibration model, which predicts the RT of benzodiazepines and illustrates the extension of the linear characteristics of the retention behavior of these drugs.

2.1 ANFIS

Over the past decade or so, significant advances have been made in two distinct technological areas: Fuzzy Logic (FL) and Neural Networks (NNs).

The fuzzy-rule based approach in modeling had been introduced by Zadeh for the first time (1965) [10]. FL and fuzzy set theory are employed to describe human thinking and reasoning in a mathematical framework. FL can represent propositions by degrees of truthfulness and falsehood and has proved to be particularly useful in artificial intelligence applications. One of the characteristics of FL is that it allows nonlinear input/output relationships to be expressed by a set of qualitative "if-then" rules [11].

In recent years, the integration of NNs and FL has given birth to A new research into neuro – fuzzy systems. These systems have the potential to combine the benefits of both fields in a single framework. Neuro – fuzzy systems eliminate the basic problem in fuzzy system design (obtaining a set of fuzzy if-then rules) by effectively using the learning capability of an ANN for automatic fuzzy if-then rule generation and parameter optimization. Figure 1 shows the basic structure of a Fuzzy Inference System (FIS). FIS implements a nonlinear mapping from its input space to the output space. This mapping is accomplished by a number of fuzzy if-then rules, each of which describes the local behavior of the mapping [12].

An adaptive network is a multilayered feed forward structure whose output behavior is determined by the value of a collection of modifiable parameters. Developing an FIS using the framework of adaptive NNs is called an ANFIS.

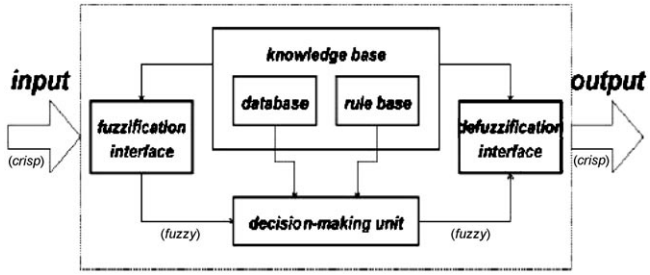

Figure 1. FIS with crisp output.

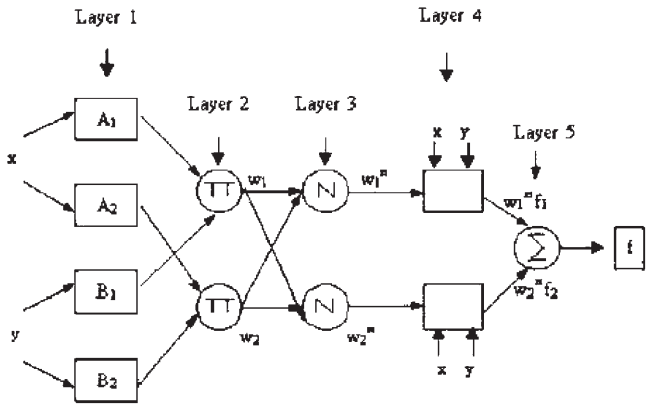

Figure 2. ANFIS architecture.

The goal of ANFIS is to find a model or mapping that will correctly associate the inputs (descriptors) with the target (activity) [13]. The general structure of the ANFIS is presented in Figure 2. Selection of the FIS is the major concern when designing an ANFIS to model a specific target system. Various types of FIS are reported in the literature (*e.g.* Mamdani and Assilian, 1975; Tsukamoto, 1979; Takagi and Sugeno, 1985). In the present work, we used the Sugeno fuzzy model [14] since the consequent part of this FIS is a linear equation and the parameters can be estimated by a simple least squares error method.

For instance, consider the FIS with two inputs x and y and one output z; for now a first order Sugeno type rule base composed of the following two rules:

$$\text{Rule 1 : If } x \text{ is } A_1 \text{ and } y \text{ is } B_1\text{; then } f_1 = p_1x + q_1y + r_1 \quad (1)$$

$$\text{Rule 2 : If } x \text{ is } A_2 \text{ and } y \text{ is } B_2\text{; then } f_2 = p_2x + q_2y + r_2 \quad (2)$$

where A_1; A_2 and B_1; B_2 are the Membership Functions (MFs) for inputs x and y, respectively; p_1; q_1; r_1, and p_2; q_2; r_2 are the parameters of the output function. The corresponding equivalent ANFIS architecture is presented in Figure 2 where nodes of the same layer have similar functions. The functioning of the ANFIS is as follows:

Layer 1: Each node in this layer generates membership grades of an input variable. The node output OP_i^1 is defined by:

$$OP_i^1 = \mu_{A_i}(x) \quad \text{for } i = 1,2 \quad \text{or} \quad (3)$$

$$OP_i^1 = \mu_{B_{i-2}}(y) \quad \text{for } i = 3,4 \quad (4)$$

where x (or y) is the input to the node; A_i (or B_{i-2}) is a fuzzy set associated with this node, characterized by the shape of the MFs in this node and can be any appropriate functions that are continuous and piecewise differentiable such as generalized bell-shaped, Gaussian, trapezoidal-shaped, and triangular-shaped functions. Assuming a generalized bell function as the MF, the output OP_i^1 can be computed as:

$$OP_i^1 = \mu_{A_i}(x) = \frac{1}{1 + \left(\frac{x - c_i}{a_i}\right)^{2b_i}} \quad (5)$$

where $\{a_i; b_i; c_i\}$ is the parameter set. As the values of the parameters change, the shape of the bell-shaped function varies with maximum equal to 1 and minimum equal to 0. Parameters in that layer are called premise parameters.

Layer 2: Every node in layer 2 is a fixed node, whose output is the product of all incoming signals. In general, any other fuzzy AND operation can be used. OP_i^2 that represents the firing strength of a rule is computed as:

$$OP_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y), \quad i = 1,2 \quad (6)$$

Layer 3: The ith node of this layer, labeled as N computes the normalized firing strengths as:

$$OP_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2}, \quad i = 1,2 \quad (7)$$

Layer 4: Every node in layer 4 is an adaptive node with a node output:

$$OP_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad (8)$$

where $\bar{w}_i$ is the output of layer 3 and $\{p_i; q_i; r_i\}$ is the parameter set of this node. Parameters in this layer are called consequent parameters [10].

Layer 5: The single node in this layer computes the overall output of the ANFIS as:

$$OP_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (9)$$

A hybrid algorithm adjusts the consequent parameters $\{p_i; q_i; r_i\}$ in a forward pass and the premise parameters $\{a_i; b_i; c_i\}$ in a backward pass [15]. In the forward pass the net-

work inputs propagate forward until layer 4, where the consequent parameters are identified by the least-squares method. In the backward pass, the error signals propagate backwards and the premise parameters are updated by a gradient descent.

Because the update rules for the premise and consequent parameters are decoupled in the hybrid-learning rule, a computational speedup may be possible by using variants of the gradient method or other optimization techniques on the premise parameters.

3 Experimental Section

3.1 Dataset

A total of 35 benzodiazepines, their active metabolites, and benzodiazepine-like substances obtained by LC-MS spectrometry taken from Ref. [2] were used as the dataset. Three compounds, norchlordiazepoxide, chlordiazepoxide, and demoxepam, which have dative bond, were removed from the dataset to prevent imprecise 3D-optimization using HyperChem software [16]. The RTs obtained by LC-MS for the 32 benzodiazepines studied in the present work are shown in Table 1. All LC-MS measurements were made with Xterra MS C-18, 150 mm × 2.1 mm, 3.5 μm PS in combination with methanol/formic acid gradient [2]. For exhaustive testing of the predictive power of the models, two Leave-Multiple-Out Cross-Validations (LMO-CV) were carried out. Here, we performed Leave-3-Out (L3O) and Leave-6-Out (L6O) cross-validations. A group of three and six compounds was randomly selected from the training set. Then each group was left out and was predicted by the model developed from the remaining observations. This procedure was carried out 200 times. In order to study the robustness of ANFIS-ANN model, we also have used Y-randomization test [17]. In this test, the RT is randomly shuffled and a new Quantitative Structure – Retention Relationship (QSRR) model is developed using the original descriptor matrix. The new QSRR model is expected to have a low multiple correlation coefficient (R^2) and multiple correlation coefficient of cross-validation (R^2_{CV}) values.

3.2 Descriptor Generation

The next step in developing a model is generation of the numerical description of the molecular structures. The numerical descriptors are responsible for encoding important features of the structure of the molecules and can be categorized as geometric, electronic, and topological properties. A total of 367 zero-, one-, two-, and three-dimensional descriptors were calculated for each compound in the dataset using the Dragon software [18]. These numbers were obtained from a total of 600 descriptors after removing the constant, zero and highly correlated ($R^2 > 0.90$) ones.

Table 1. Experimental and ANFIS-ANN calculated values of LC-MS RTs of benzodiazepines studied in this work.

No.	Compound	Experimental RT	Calculated RT
1[a]	Zolpidem	5.35	8.27
2[b]	Aminoclonazepam	5.64	12.27
3[a]	Acetamidonitrazepam	6.77	10.37
4[c]	7-Aminoflunitrazepam	7.90	6.55
5[a]	Desmethylmedazepam	9.46	12.22
6[a]	Flurazepam	9.91	9.33
7[a]	Loprazolam	9.91	9.67
8[b]	Acetamidoclonazepam	10.04	7.88
9[a]	Midazolam	10.57	10.03
10[c]	OH-midazolam	15.78	11.91
11[a]	Bromazepam	17.16	20.69
12[a]	Zopiclone	17.47	21.40
13[a]	OH-Bromazepam	18.33	25.11
14[b]	N-Desmethylflunitrazepam	19.15	20.41
15[a]	Nitrazepam	20.88	21.56
16[c]	Desmethylclobazam	21.37	22.38
17[a]	Clonazepam	22.39	20.48
18[a]	OH-triazolam	22.72	30.46
19[a]	Flunitrazepam	22.96	20.69
20[b]	α-OH-alprazolam	23.83	26.16
21[a]	Clobazam	25.04	22.42
22[c]	Alprazolam	27.16	25.10
23[a]	OH-ethylflurazepam	27.44	28.13
24[a]	Oxazepam	27.45	24.68
25[a]	Triazolam	27.93	29.87
26[b]	Lorazepam	28.07	27.28
27[a]	Brotizolam	29.81	31.38
28[c]	Desalkylflurazepam	29.95	23.06
29[a]	Temazepam	29.97	30.26
30[a]	Lormetazepam	32.05	30.53
31[a]	Nordazepam	32.35	26.89
32[b]	Diazepam	34.82	30.62

[a]Training set.
[b]Test set.
[c]Validation set.

3.3 Regression Analysis

The stepwise MLR procedure was used for model generation. The stepwise addition method implemented in the software package of SPSS [19] was used for choosing the descriptors contributing to the RT of benzodiazepines. The more suitable models obtained in each stage, were compared and among them, the best MLR model was chosen for further evaluation. The best MLR model consists of three descriptors of 3D-MoRSE-signal 06/weighted by atomic polarizabilities, Mor06p, R maximal autocorrelation of lag 4/weighted by atomic Sanderson electronegativities, R4e+ and 2nd component accessibility directional WHIM index/weighted by atomic masses, E2m (Table 2). The main goals of generating the MLR model were developing a linear model for considering the linear characteristic of the retention of benzodiazepines as well as choosing a set of suitable descriptors as inputs for developing the ANN.

Table 2. The best MLR model for the prediction of benzodiazepines RT together with the mean effects of the descriptors appearing in the model.

Descriptors	Notation[a]	Regression coefficient	Mean effect[b]
(1) 3D-MoRSE-signal 06/weighted by atomic polarizabilities	Mor06p	14.873 (±2.817)	12.186
(2) R maximal autocorrelation of lag 4/weighted by atomic Sanderson electronegativities	R4e+	256.795 (±67.833)	14.204
(3) 2nd component accessibility directional WHIM index/weighted by atomic masses	E2m	12.282 (±8.228)	5.665
Constant		−11.101 (±5.117)	

[a] The statistics for the model are as follows: $n=26$, $q^2=0.646$, $RMSE_{LOO}=5.828$, $F=19.533$.
[b] Mean effect of a descriptor is the product of its mean and regression coefficient in the MLR model.

3.4 ANFIS

Sequential forward search was used as a tool for input selection by ANFIS modeling. The m-function of fuzzy toolbox of MATLAB 6.5. was used for this purpose. The algorithm is based on selecting the best descriptor, which minimizes the Standard Error of Training (SE_T) and test sets. Repeatedly, next descriptor will be selected by adding it to the previous one. The number of ANFIS modeling processes for selecting the N inputs (here 3) and M candidates (here 367) with sequential forward search can be calculated as:

$$\text{No. of ANFIS models} = \frac{((2\times M)-N+1)\times N}{2} \qquad (10)$$

3.5 ANN

A detailed description of the theory behind a neural network has been adequately described elsewhere [20, 21]. ANN was developed as a nonlinear mapping technique for both MLR and ANFIS feature selection methods. The feed forward Back Propagation Artificial Neural Networks (BP-ANN) used in this work was written in MATLAB version 6.5 as an m-file. The output layer represents the RT of benzodiazepines. In this investigation, the sigmoid transfer function was used as the transfer function. The initial values of biases were set to be one. The initial weights were chosen randomly between −0.3 and +0.3, and optimized based on the Levenberg – Marquardt optimization as a fast function in updating the weights of the network [22, 23].

Before training, the inputs were normalized between −2 and 2 and the output between 0.2 and 0.8. The network parameters such as the number of neurons in the hidden layer, learning rate and momentum were optimized based on obtaining the minimum Standard Error of Training (SET) and Standard Error of Prediction (SE_P) [21, 24]. The number of neurons of the hidden layer with the minimum value of SE was selected as the optimum number. Then learning rate and momentum were optimized in a similar way. Also, we used LMO-CV for considering the consistency of the final generated model and also comparing the results of ANFIS-ANN with MLR-ANN for which the variable selection method is linear.

The SZW, presented by Nord and Jacobsson [25], was used to illustrate the contribution of each variable to the retention of benzodiazepines. This method is based on SWZ of the connection between the input variables and the first hidden layer of the generated ANN model. A program is written in MATLAB version 6.5. as an m-file in a way that zeroing the weights for each input variable results in a measure of the importance of that variable. The measure of variable influence used in this work is Rmdiff (Eq. 11). This parameter is the difference between root mean standard error of predictions, $RMSE_P$, when weights in the ANN model are zeroed to exclude one variable and the $RMSE_P$ of the complete model [26, 27].

$$\text{Rmdiff} = \sqrt{\frac{1}{N}\sum_k (y_k^{\text{Zeroed}} - t_k)^2} - \sqrt{\frac{1}{N}\sum_k (y_k^{\text{fullNN}} - t_k)^2} \qquad (11)$$

In this equation, y_k^{Zeroed} is the predicted ANN response when one variable is excluded, y_k^{fullNN} is the ANN response for the full ANN model, t_k is the experimental value and N is the number of samples used in the calculation of Rmdiff. A large value of Rmdiff shows a significant contribution from that variable [26].

4 Results and Discussion

Our main goal in this study was comparing the two feature selection methods of stepwise regression adopted for MLR and ANFIS to model the retention behavior of benzodiazepines. In fact, we intended to study the differences of the two methods of MLR and ANFIS as a tool for selecting suitable variables. To reach this goal we have considered the following steps:

4.1 Multiple Regression Analysis as Feature-Selection and Calibration Model

Linear models were formed by a stepwise addition of terms. A deletion process was then employed where each

Table 3. Statistics of L3O and L6O cross-validations for the comparison of MLR, MLR-ANN, and ANFIS-ANN methods.

	MLR[a]				MLR-ANN[b]				ANFIS-ANN[b]			
	R^2	RMSE	R^2_{CV}	$RMSE_{CV}$	R^2	RMSE	R^2_{CV}	$RMSE_{CV}$	R^2	RMSE	R^2_{CV}	$RMSE_{CV}$
L3O	0.6221	5.531	0.616	5.891	0.716	4.903	0.674	6.390	0.847	2.940	0.792	4.680
L6O	0.618	5.548	0.589	5.711	0.734	4.731	0.632	6.481	0.849	2.966	0.771	4.792

[a]The number of compounds in the training and test sets are 29 and 3 for L3O and 26 and 6 for L6O, respectively.
[b]The number of compounds in the training, test, and validation sets are 24, 5, 3 for L3O and 21, 5, and 6 for L6O, respectively.

Table 4. Architecture and specifications for hybrid methods of MLR-ANN and ANFIS-ANN.

	MLR-ANN	ANFIS-ANN
No. of nodes in the input layer	3	3
No. of nodes in the hidden layer	4	4
No. of nodes in output layer	1	1
Learning rate	0.5	0.1
Momentum	0.5	0.9
Transfer function	Sigmoid	Sigmoid
Update weight function	Levenberg–Marquardt ($\mu=0.3$, $\lambda=10$)	Levenberg–Marquardt ($\mu=0.1$, $\lambda=10$)

variable in the model was held out in turn and a model was generated by using the remaining parameters [28]. A correlation matrix was formed for all 367 descriptors calculated using Dragon software [18]. Inspection of this matrix did not show a considerable correlation ($R^2>0.90$) between the descriptors. In the first step, after sorting the dataset (Table 1) based on the RT values, training and prediction sets in a ratio of 4:1 were chosen in a way that the prediction set adequately represents the training set. Then, a stepwise selection procedure implemented in SPSS software [19] was used to select the descriptors using the training set ($n=26$ compounds). Three descriptors of Mor06p, R4e+, and E2m were chosen as a group of parameters producing the best MLR model. The specifications for the best selected MLR model are shown in Table 2. Also, the mean effect for each descriptor is included in this table. The statistical parameters of L3O and L6O cross-validation for generated MLR model are gathered in Table 3. The poor results show the disability of this linear model in predicting the RT of benzodiazepines. It could be due to a nonlinear relationship between the RT and the MLR descriptors.

We have applied L3O and L6O cross-validations for the Partial Least Square (PLS) method [29] on a 32×367 matrix of descriptors as X matrix and 32×1 matrix of RTs as Y matrix. For this purpose, we randomly left three or six compounds out in each step, generated a PLS model based on remaining data and predicted the left out compounds. We have performed the L3O and L6O cross-validations for 200 runs. The low values for $R^2_{L3O}=0.663$ and $R^2_{L6O}=0.615$ using nine latent variables prove the nonlinearity of the system under study. This is in agreement with the conclusion obtained from the stepwise regression approach adopted for MLR with poor statistical results and indicates a nonlinear characteristic for the retention mechanism of benzodiazepines in LC. This conclusion prompted us to develop an ANN as a nonlinear model for predicting the RT of a variety of benzodiazepines in LC.

4.2 The Hybrid Chemometric Method of MLR-ANN

There were two purposes for developing the MLR model: (1) as a subjective feature selection tool for choosing inputs for the neural network and (2) providing the possibility for comparison of the ability of linear and nonlinear models in predicting the benzodiazepines RT. For the sake of comparison, the descriptors used in the MLR model should be the same as input parameters for generating the network. Therefore, a BP-ANN was generated by using the three descriptors appearing in the MLR model as its inputs. Table 4 shows the architecture and specifications of the optimized ANNs.

To check the reliability of the proposed model we have used LMO-CV. Based on this technique, a number of modified datasets were created by deleting in each step a small group of objects (here three and six objects) and then the model was evaluated by measuring its accuracy in predicting the responses of the deleted group (the ones that have not been utilized in the development of the model). The results of L3O and L6O for the two methods of MLR and ANN are reported in Table 3. The poor statistics of R^2, Root Mean Squared Error (RMSE), R^2_{cv}, and Root Mean Squared Error of Cross-Validation ($RMSE_{CV}$) for different datasets of L3O and L6O reveals the disability of both models of MLR and MLR-ANN in predicting the LC-MS RT of benzodiazepines. As it is clear, MLR-ANN results show a little improvement in the predictive ability of the hybrid method of MLR-ANN over that of MLR.

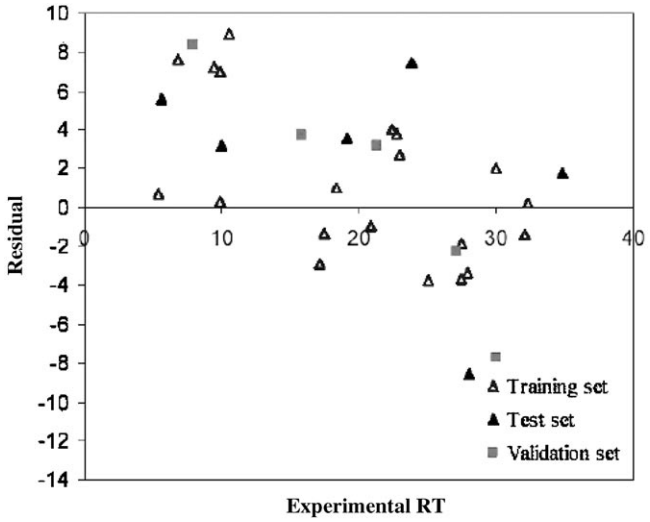

Figure 3. Plot of residuals *versus* experimental values of RT for the MLR-ANN hybrid model.

To investigate the existence of a systematic error in the generated ANN model, the residuals of ANN predicted values of C-18 RTs were plotted against the experimental values in Figure 3. The propagation of residuals on one side of zero line for low values of the RTs indicates that a systematic error exists in the development of neural network. This propagation also existed in MLR residual plot (it has not been shown here). However, ANN model with the MLR descriptors as its inputs cannot solve the problem in a nonlinear modeling. This may be due to the fact that the descriptors of the MLR with a linear characteristic are not a good candidate for the inputs of the nonlinear method of ANN. Therefore, using these descriptors in building a nonlinear model may result in a special systematic trend. This prompted us to use ANFIS as a nonlinear feature selection method, hoping to improve the results. It is noteworthy that ANFIS is capable to adapt with linear and nonlinear systems but choosing ANFIS in the present work was based on its adaptation with nonlinear phenomena.

4.3 Application of ANFIS as Feature Selection Technique

In the present contribution, the sequential forward search was used for choosing ANN inputs using ANFIS method. For running ANFIS program, after sorting the dataset based on the RT values (Table 1), training and test sets in a ratio of 4:1 were chosen (the same as for MLR) in a way that the prediction set adequately represent the training set. Three out of 367 descriptors were selected based on two Gaussian MFs and three iterations. Table 5 shows the specifications of the ANFIS model in choosing the descriptors based on the sequential forward search. Three descriptors of a number of functional groups of R−Co−N< and >N−C=N (N-072), Mor06p, and Radial Distribution Function-1.0/weighted by atomic van der Waals volumes (RDF010v) were used as inputs for developing the ANN model. This paves the way for comparison between the two feature selection methods of stepwise regression adopted for MLR and ANFIS, as indicated in the next step.

Table 5. Specifications of the ANFIS for input selection.

Type	Sugeno
AND method	Product
Implication method	Product
Number of memberships functions for inputs	2
Type of MFs	Gaussian
Number of rules	8
Output MF	Constant
Number of neuro – fuzzy epochs	3
Type of ANFIS search for input selection	Sequential forward search

4.4 The Hybrid Chemometric Method of ANFIS-ANN

Three descriptors selected by using ANFIS were used as inputs of BP-ANN. Table 4 shows the architecture and specifications of the optimized ANFIS-ANN model. We employed a two-step validation protocol. The ANFIS-ANN model is first validated internally using the dataset. The dataset was divided into training (calibration), test, and validation sets after sorting based on the RT values. The training set consisted of 21 molecules and the test and validation sets, consisted of six and five molecules, respectively. The test and the validation sets adequately represent the training set. The training set was used for model generation. The test set was applied to prevent the overfitting of the network. However, the validation set in which its molecules have no role in model building was used for the evaluation of the predictive ability of the trained network. Correlation coefficient (R^2) values of 0.829, 0.823, and 0.890 were obtained for the training, test, and validation sets, respectively. Also, RMSEs of 3.282, 3.510, and 3.730, respectively, were obtained for these sets. The ANFIS-ANN calculated values of RT for the training, test, and validation sets are presented in Table 1.

Figure 4 shows the plot of the ANFIS-ANN predicted *versus* the experimental values for ANFIS-ANN. The agreement between the predicted and observed RTs using the ANFIS-ANN shows the ability of the nonlinear feature selection/nonlinear modeling in predicting the RT of benzodiazepines.

The residuals of the calculated values of RT are plotted against the experimental values in Figure 5 for the sets of training, test, and validation. The propagation of the residuals on both sides of zero line indicates that there is no systematic error in the generated model. It is noteworthy that a similar plot for the MLR-ANN model shows some con-

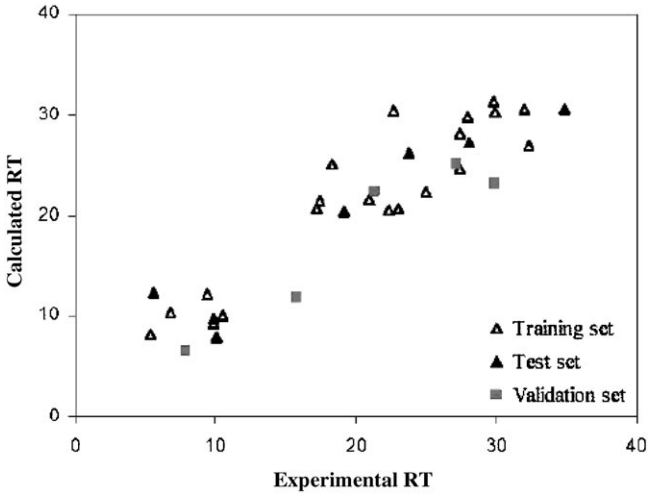

Figure 4. Plot of the ANFIS-ANN calculated RTs against the experimental values for all benzodiazepines studied in this work.

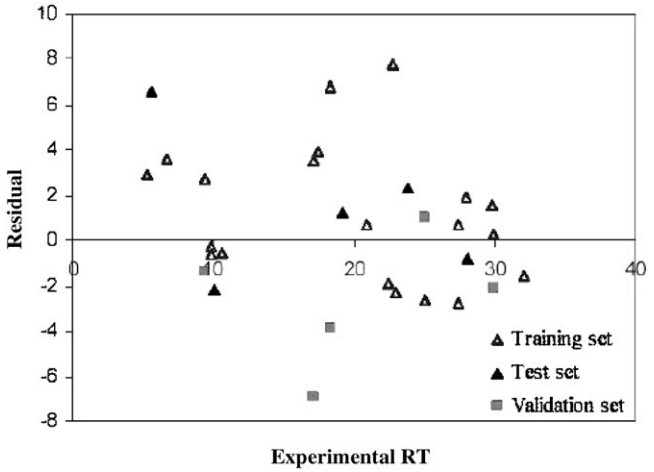

Figure 5. Plot of residuals *versus* experimental values of RT for the ANFIS-ANN hybrid model.

tradictions. Most of the MLR-ANN calculated RTs for the compounds eluted before around 15 min seem to be overestimated. This systematic trend does not exist for the generated ANFIS-ANN model.

As a second step of validation protocol, the L3O and L6O cross-validations were performed on ANFIS-ANN model (Table 3). Comparing the statistics of this model with the previous ones of MLR-ANN reveals the superiority of ANFIS as a nonlinear feature selection method over the stepwise regression MLR as the linear one in predicting the RT of benzodiazepines. However, both methods of MLR-ANN and ANFIS-ANN have the same optimized architecture from the point of the number of hidden layer nodes but their learning parameters are different. Although, learning rate, momentum, and weight update function values have some role in optimization but this role is

Table 6. R^2 and R^2_{L6O} values for ANFIS-ANN model after several Y-randomization tests.

Model	R^2	R^2_{L6O}
1	0.128	0.011
2	0.144	0.121
3	0.013	0.002
4	0.214	0.101
5	0.012	0.028
6	0.238	0.144
7	0.112	0.008
8	0.129	0.112
9	0.116	0.013
10	0.321	0.204

a minor one. The most important factor is the type of the input descriptors. In fact, this research shows the power of ANFIS in choosing the most suitable parameters. In other words, the descriptors play the major role. However, their number is identical for both methods but their effectiveness is different.

The ANFIS-ANN model was further evaluated by applying the Y-randomization test. Several random shuffles of the Y (RT) were chosen and the feature selection and modeling process were performed for all cases. The results are shown in Table 6. The low values for R^2 and R^2_{L6O} show that the good statistical results in ANFIS-ANN model are not due to a chance correlation or structural dependency of the training set [30].

The RDF descriptor of RDF010v is a geometrical descriptor weighted by atomic van der Waals volumes and is related to the three-dimensional structure of compounds. This parameter can show the importance of size and the direction of compounds when reacting with the stationary phase in the column. Therefore, it indicates the role of dispersion interactions on the retention mechanism of benzodiazepines on C-18 stationary phase [31]. Mor06p as a 3D-MoRSE represents the 3D geometry of compounds and is concerned with the electronic diffraction of compounds especially when it is weighted by polarizability. This effect may show the interaction between polar sites of benzodiazepines with polar mobile phase of methanol/formic acid in LC-MS. This factor reveals the importance of electrostatic interaction between benzodiazepines passing through the nonpolar column and polar mobile phase. N-072 shows the number of functional groups of R–Co–N < and > N–C=N in benzodiazepines. Due to the polar nature of these functional groups, it may result in the formation of hydrogen bond between benzodiazepine molecule and polar mobile phase of methanol/formic acid in LC-MS. Therefore, the interactions of dispersion, electrostatic, and hydrogen bond can be responsible for the retention of these compounds in the retention process of LC-MS.

In order to show the extent of contribution for each descriptor to the RT of benzodiazepines, the parameter of

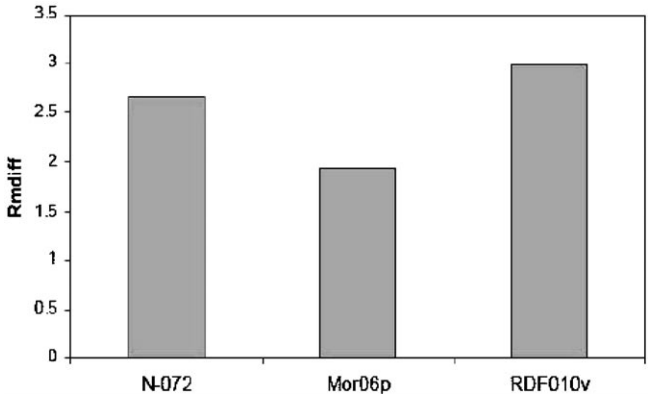

Figure 6. SZW as a sensitivity analysis method for ANFIS-ANN descriptors.

Rmdiff (Eq. 11) was calculated as a measure of variable influence and the results are depicted in Figure 6 for the three descriptors. It can be seen from this figure that RDF010v shows the largest contribution to the retention. This reveals that the dispersion interactions play a major role in the retention behavior of benzodiazepines on C-18 column. We have also plotted the variation of the ANFIS selected descriptors against LC-MS RTs of different benzodiazepines (Figure 7). It is clear from these plots that among the three descriptors only RDF010v shows an exponential nonlinear trend against RT. This behavior in agreement with the results of SZW analysis shows the role of this descriptor in nonlinearity of the retention process and its high contribution to this phenomenon.

5 Conclusions

The main drawbacks of ANNs are twofold: (1) the NNs are unable to choose their inputs. Nowadays, the number of structural parameters for a molecule is very large and increasing (more than 1497 descriptors can be calculated using Dragon software, version 3.0). Therefore, choosing a few effective ones among a large number is a challenging process and (2) the purpose of developing ANN models is mainly due to their prediction abilities, while the chemists are interested in developing models having both predictive and interpretive characteristics. One way of a model being interpretive is having interpretable inputs and finding the extent of contribution of each descriptor to the quality of interest.

Keeping both points in mind, we have successfully applied ANFIS for selecting the best parameters for modeling LC-MS RT of benzodiazepines. It is shown that the 3-4-1 ANFIS-ANN model is superior over the MLR-ANN

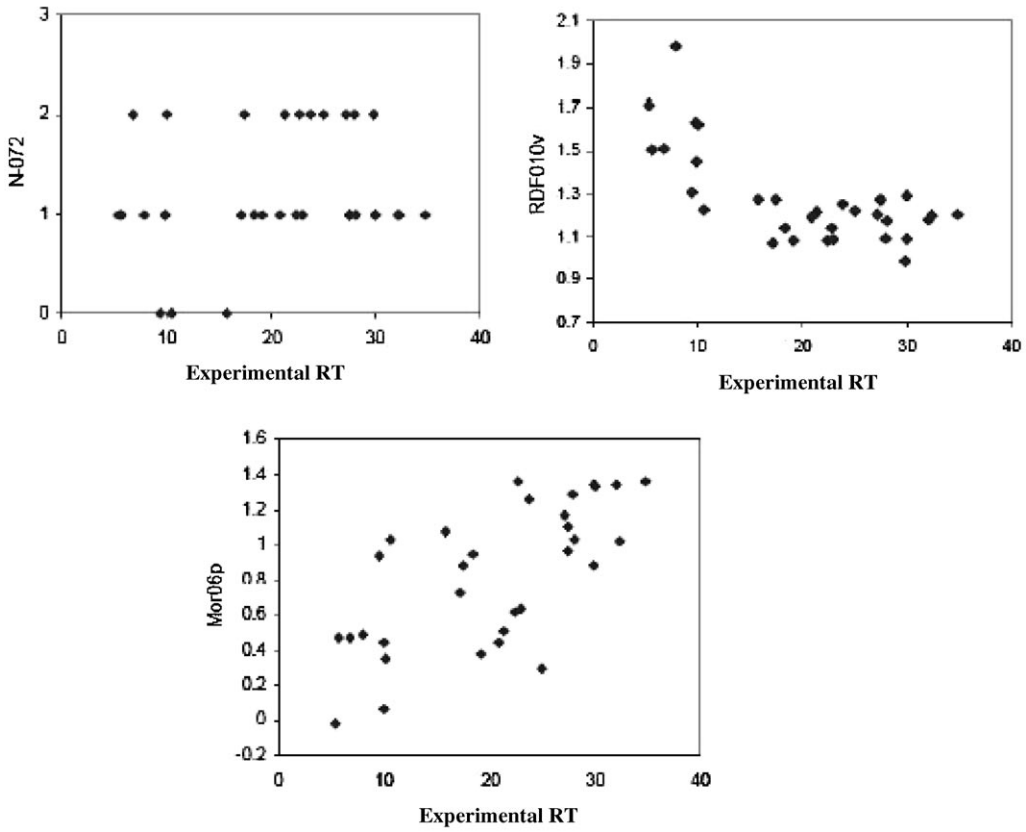

Figure 7. Plots of variations of the three ANFIS descriptors against the experimental values of RT of benzodiazepines.

model. Superiority of the former model suggests a nonlinear characteristic for the retention behavior of LC-MS data. The appearance of three descriptors of N-072, Mor06p, and RDF010v revealed the importance of dispersion, electrostatic, and hydrogen bond interactions between benzodiazepines and C-18 stationary and polar mobile phases, respectively. In order to assess the importance of each descriptor, SWZ was applied. SZW shows the largest contribution for RDF010v, which means that the dispersion interactions play a major role in the retention behavior of benzodiazepines on C-18 column.

References

[1] O. H. Drummer, *Forensic Sci. Rev.* **2002**, *14*, 1 – 2.
[2] B. E. Smink, J. E. Brandsma, A. Dijkhuizen, K. J. Lusthof, J. J. de Gier, A. C. G. Egberts, D. R. A. Uges, *J. Chromatogr. B* **2004**, *811*, 13 – 20.
[3] F. Barbone, A. D. McMahon, P. G. Davey, A. D. Morris, I. C. Reid, D. G. McDevitt, T. M. MacDonald, *Lancet* **1998**, *352*, 1331 – 1336.
[4] M. M. Ariffin, R. A. Anderson, *J. Chromatogr. B* **2006**, *842*, 91 – 97.
[5] B. M. Chen, Y. Z. Liang, X. Chen, S. G. Liu, F. L. Deng, P. Zhou, *J. Pharm. Biomed. Anal*. **2006**, *42*, 480 – 487.
[6] M. J. Bogusz, R.-D. Maier, K.-D. Kruuger, W. Fruuchtnicht, *J. Chromatogr. B* **1998**, *713*, 361 – 369.
[7] M. R. Moeller, T. Kraemer, *Ther. Drug Monit*. **2002**, *24*, 210 – 221.
[8] M. Jalali-Heravi, A. Kyani, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1328 – 1335.
[9] T. Hancock, R. Put, D. Coomans, Y. V. Heyden, Y. Everingham, *Chemom. Intell. Lab. Syst.* **2005**, *76*, 185 – 196.
[10] L. A. Zadeh, *Inf. Control*. **1965**, *8*, 338 – 353.
[11] G. J. Klir, B. Yuan, *Fuzzy Sets and Fuzzy Logic, Theory and Applications*, Prentice-Hall, Inc., New Jersey **1995**.
[12] P. C. Nayak, K. P. Sudheer, D. M. Rangan, K. S. Ramasastri, *J. Hydrol.* **2004**, *291*, 52 – 66.
[13] Y. L. Loukas, *J. Med. Chem*. **2001**, *44*, 2772 – 2783.
[14] T. Takagi, M. Sugeno, *IEEE Trans. Syst. Man Cybern.* **1985**, *15*, 116 – 132.
[15] J.-S. R. Jang, C.-T. Sun, *Proc. IEEE* **1995**, *83*, 378 – 406.
[16] HyperChem, *Molecular Modeling System*, Hyper Cube, Inc. and Auto esk, Inc. **1993**.
[17] A. Tropsha, P. Gramatica, V. K. Gombar, *QSAR Comb. Sci.* **2002**, *22*, 69 – 77.
[18] R. Todeschini, V. Consonni, A. Mauri, M. Pavan, via Pisani, 13 – 20124 Milano, Italy, Dragon Software version 3.0, 2003.
[19] SPSS for Windows, version 10.05, Standard version, SPSS Inc. 1999.
[20] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, VCH, Weinheim **1999**.
[21] M. Jalali-Heravi, Z. Garkani-Nejad, *J. Chromatogr. A* **2001**, *927*, 211 – 218.
[22] A. Guven, S. Kara, *Exp. Syst. Appl.* **2006**, *31*, 199 – 205.
[23] M. T. Hugan, M. B. Menhaj, *IEEE Trans. Neural Netw.* **1994**, *5*, 989 – 993.
[24] T. B. Blank, S. T. Brown, *Anal. Chem*. **1993**, *65*, 3081 – 3089.
[25] L. I. Nord, S. P. Jacobsson, *Chemom. Intell. Lab. Syst.* **1998**, *44*, 153 – 160.
[26] F. O. Andersson, M. Aberg, S. P. Jacobsson, *Chemom. Intell. Lab. Syst.* **2000**, *51*, 61 – 72.
[27] S. Papadokonstantakis, A. Lygeros, S. P. Jacobsson, *Neural Networks* **2006**, *19*, 500 – 513.
[28] N. Draper, H. Smith, *Applied Regression Analysis*, Wiley-Interscience, New York, 2nd Edn. **1981**, p. 307.
[29] S. Wold, M. Sjostorm, L. Eriksson, *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109 – 130.
[30] J. G. Topliss, R. P. Edwards, *J. Med. Chem.* **1979**, *22*, 1238 – 1244.
[31] M. Jalali-Heravi, Z. Garkani-Nejad, A. Kyani, *QSAR Comb. Sci.*, in press. Doi: 10.1002/qsar200510205.

