Combining classifiers of pesticides toxicity through a neuro-fuzzy approach

Emilio Benfenati[1], Paolo Mazzatorta[1], Daniel Neagu[2], and Giuseppina Gini[2]

[1] Istituto di Ricerche Farmacologiche "Mario Negri" Milano,
Via Eritrea, 62, 20157 Milano, Italy
{Benfenati, Mazzatorta}@ marionegri.it
[2] Dipartimento di Elettronica e Informazione, Politecnico di Milano,
Piazza L. da Vinci 32, 20133 Milano, Italy
Neagu@fusberta.elet.polimi.it, Gini@elet.polimi.it
http://airlab.elet.polimi.it/imagetox

Abstract. The increasing amount and complexity of data in toxicity prediction calls for new approaches based on hybrid intelligent methods for mining the data. This focus is required even more in the context of increasing number of different classifiers applied in toxicity prediction. Consequently, there exist a need to develop tools to integrate various approaches. The goal of this research is to apply neuro-fuzzy networks to provide an improvement in combining the results of five classifiers applied in toxicity of pesticides. Nevertheless, fuzzy rules extracted from the trained developed networks can be used to perform useful comparisons between the performances of the involved classifiers. Our results suggest that the neuro-fuzzy approach of combining classifiers has the potential to significantly improve common classification methods for the use in toxicity of pesticides characterization, and knowledge discovery.

1 Introduction

Quantitative structure–activity relationships (QSARs) correlate chemical structure to a wide variety of physical, chemical, biological (including biomedical, toxicological, ecotoxicological) and technological (glass transition temperatures of polymers, critical micelle concentrations of surfactants, rubber vulcanization rates) properties. Suitable correlations, once established and validated, can be used to predict properties for compounds as yet unmeasured or even unknown.

Classification systems for QSAR studies are quite usual for carcinogenicity [9], because in this case carcinogenicity classes are defined by regulatory bodies such as IARC and EPA. For ecotoxicity, most of the QSAR models are regressions, referring to the dose giving the toxic effect in 50% of the animals (for instance LC_{50}: lethal concentration for 50% of the test animals). This dose is a continuous value and regression seems the most appropriate algorithm. However, classification affords some advantages. Indeed, i) the regulatory values are indicated as toxicity classes and ii) classification can allow a better management of noisy data. For this reason we investigated classification in the past [7], [8], [9] and also in this study. No general

rule exists to define an approach suitable to solve a specific classification problem. In several cases, a selection of descriptors is the only essential condition to develop a general system. The next step consists in defining the best computational method to develop robust structure–activity models.

Artificial neural networks (ANNs) represent an excellent tool that have been used to develop a wide range of real-world applications, especially when traditional solving methods fail [3]. They exhibit advantages such as ideal learning ability from data, classification capabilities and generalization, computationally fastness once trained due to parallel processing, and noise tolerance. The major shortcoming of neural networks is represented by their low degree of human comprehensibility. More transparency is offered by fuzzy neural networks FNN [14], [16], [18], which represent a paradigm combining the comprehensibility and capabilities of fuzzy reasoning to handle uncertainty, and the capabilities to learn from examples.

The paper is organized as follows. Section 2 briefly presents the aspects of data preparation, based on chemical descriptors, some of the most common classification techniques and shows how they behave for toxicology modeling, with a emphasis to pesticides task. Section 3 proposes the neuro-fuzzy approach in order to manage the integration of all the studied classifiers, based on the structure developed as FNN Implicit Knowledge Module (IKM) of the hybrid intelligent system NIKE (Neural explicit&Implicit Knowledge inference system [17]). Preliminary results indicate that combination of several classifiers may lead to the improved performance [5], [11], [12]. The extracted fuzzy rules give new insights about the applicability domain of the implied classifiers. Conclusions of the paper are summarized in the last section.

2 Materials and Methods

2.1 Data set

For this paper a data set constituted of 57 common organophosphorous compounds has been investigated. The main objective is to propose a good benchmark for the classification studies developed in this area. The toxicity values are the result of a wide bibliographic research mainly from "*the Pesticide Manual*", ECOTOX database system, RTECS and HSDB [1]. An important problem that we faced is connected with the variability that the toxicity data presents [2]. Indeed, it is possible to find different fonts showing for the same compound and the same end–point LC_{50} different for about two orders of magnitude. Such variability is due to different factors, as the different individual reactions of organisms tested, the different laboratory procedures, or is due to different experimental conditions or accidental errors.

The toxicity value was expressed using the form Log_{10} ($1/LC_{50}$). Then the values were scaled in the interval [-1..1]. Four classes were defined: Class 1 [-1..-0.5), Class 2 [-0.5..0), Class 3 [0..0.5), Class 4 [0.5..1] (Table 2).

2.2 Descriptors

A set of about 150 descriptors were calculated by different software: Hyperchem 5.0[1], CODESSA 2.2.1[2], Pallas 2.1[3]. They are split into six categories: Constitutional (34 descriptors), Geometrical (14), Topological (38), Electrostatic (57), Quantum–chemicals (6), and Physico–chemical (4). In order to obtain a good model, a selection of the variables, which better describe the molecules, is necessary. There is the risk that some descriptors does not add information, and increase the noise, making more complex the result analysis. Furthermore, using a relatively low number of variables, the risk of overfitting is reduced. The descriptors selection (table 1) was obtained by Principal Components Analysis (PCA), using SCAN[4]:

Table 1. Names of the chemical descriptors involved in the classification task.

	Cat.	Cod.
Moment of inertia A	G	D1
Relative number of N atoms	C	D2
Binding energy (Kcal/mol)	Q	D3
DPSA-3 Difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]	E	D4
Max partial charge (Qmax) [Zefirov's PC]	E	D5
ZX Shadow / ZX Rectangle	G	D6
Number of atoms	C	D7
Moment of inertia C	G	D8
PNSA-3 Atomic charge weighted PNSA [Zefirov's PC]	E	D9
HOMO (eV)	E	D10
LUMO (eV)	Q	D11
Kier&Hall index (order 3)	T	D12

2.3 Classification algorithms

The classification algorithms used for this work are five: LDA (Linear Discriminant Analysis), RDA (Regularized Discriminant Analysis), SIMCA (Soft Independent Modeling of Class Analogy), KNN (K Nearest Neighbors classification), CART (Classification And Regression Tree). The first four are parametric statistical systems based on the Fisher's discriminant analysis, the fifth and sixth are not parametrical statistical methods, the last one is a classification tree.

LDA: the Fischer's linear discrimination is an empirical method based on p–dimensional vectors of attributes. Thus the separation between classes occurs by an hyperplane, which divides the p–dimensional space of attributes.

RDA: The variations introduced in this model have the aim to obviate the principal problems that afflict both the linear and quadratic discrimination. The regulation more efficient was carried out by Friedman, who proposed a compromise between the two previous techniques using a biparametrical method for the estimation (λ and γ).

[1] Hypercube Inc., Gainsville, Florida, USA
[2] SemiChem Inc., Shawnee, Kansas, USA
[3] CompuDrug; Budapest, Hungary
[4] SCAN (Software for Chemometric Analysis) v.1.1, from Minitab: http://www.minitab.com

SIMCA: the model is one of the first used in chemometry for modeling classes and, contrarily to the techniques before described, is not parametrical. The idea is to consider separately each class and to look for a representation using the principal components. An object is assigned to a class on the basis of the residual distance, rsd^2, that it has from the model which represent the class itself:

$$r_{igj}^2 = \left(\hat{x}_{igj} - x_{igj}\right)^2, \; rsd^2 = \frac{\sum_j r_{igj}^2}{(p - M_j)} \qquad (\mathbf{1})$$

where x_{igj} = co–ordinates of the object's projections on the inner space of the mathematical model for the class, x_{igj} = object's co–ordinates, p=number of variables, M_j = number of the principal components significant for the j class.

KNN: this technique classifies each record in a data set based on a combination of the classes of the *k* record(s) most similar to it in a historical data set (where $k = 1$).

CART is a tree–shaped structure that represents sets of decisions. These decisions generate rules for the classification of a data set. CART provides a set of rules that can be applied to a new (unclassified) data set to predict which records will have a given outcome. It segments a data set by creating two–way splits.

The classification obtained using these algorithms is shown in Table 2.

2.4 Validation

The more common methods for validation are: i) Leave–one–out (LOO); ii) Leave–more–out (LMO); iii) Train & Test; iv) Bootstrap. We used LOO, since it is considered the best working on data set of small dimension [10]. According to LOO, given *n* objects, *n* models are computed. For each model, the training set consists of *n*–1 objects and the evaluation set consists of the object left. To estimate the predictive ability, we considered the gap between the experimental (fitting) and the predicted value (cross–validation) for the *n* objects left, one by one, out from the model.

Table 2. True class and class assigned by the algorithms for each compound[5].

	True Class	CART	LDA	KNN	SIMCA	RDA
Anilofos	2	2	2	1	2	2
Chlorpyrifos	1	2	2	1	2	2
Chlorpyryfos-methyl	2	2	2	1	2	2
Isazofos	1	1	1	2	1	1
Phosalone	2	2	2	2	2	2
Profenofos	1	2	2	1	2	2
Prothiofos	2	2	2	2	2	2
Azamethiphos	2	2	2	1	4	2
Azinphos methyl	1	1	1	2	1	1
Diazinon	3	3	1	1	4	1
Phosmet	2	2	2	1	2	2
Pirimiphos ethyl	1	1	1	1	1	1
Pirimiphos methyl	2	3	1	2	1	1

[5] The 40 molecules with a blank background were used to train the neuro-fuzzy classifier.

Pyrazophos	2	2	1	4	2	1
Quinalphos	1	1	1	2	1	1
Azinphos-ethyl	1	1	1	1	2	1
Etrimfos	1	1	1	3	3	1
Fosthiazate	4	2	2	2	4	2
Methidathion	1	1	1	1	1	1
Piperophos	3	3	3	2	2	3
Tebupirimfos	4	1	1	3	4	1
Triazophos	1	1	1	2	1	1
Dichlorvos	2	4	2	2	2	2
Disulfoton	3	3	3	1	3	3
Ethephon	4	4	4	4	4	4
Fenamiphos	1	1	3	2	1	1
Fenthion	2	2	3	2	2	3
Fonofos	1	1	3	2	1	3
Glyphosate	4	4	4	4	4	4
Isofenphos (isophenphos)	3	3	3	1	3	3
Methamidophos	4	4	4	3	4	4
Omethoate	3	3	3	3	3	3
Oxydemeton-methyl	3	3	3	3	3	3
Parathion ethyl (parathion)	2	2	2	3	1	3
Parathion methyl	3	3	3	3	3	3
Phoxim	2	2	1	1	1	1
Sulfotep	1	1	3	2	2	2
Tribufos	2	2	2	2	2	2
Trichlorfon	2	2	2	1	2	4
Acephate	4	4	1	3	4	4
Cadusafos	2	2	3	3	2	2
Chlorethoxyfos	2	2	2	3	2	2
Demeton-S-methyl	3	3	3	3	3	3
Dimethoate	3	3	1	1	3	3
Edifenphos	2	2	3	1	2	2
EPN	2	2	2	2	2	2
Ethion	2	2	2	2	2	2
Ethoprophos	3	3	3	2	2	3
Fenitrothion	3	2	3	3	3	3
Formothion	3	3	2	3	3	3
Methacrifos	2	2	2	2	2	3
Phorate	1	1	3	2	1	3
Propetamphos	3	3	3	4	2	3
Sulprofos	3	3	3	2	3	3
Temephos	3	3	2	1	3	2
Terbufos	1	1	3	2	3	3
Thiometon	3	3	3	3	3	3

3.1 The neuro-fuzzy combination of the classifiers

3.2 Motivations and architecture

Combining multiple classifiers could be considered as a direction for the development of highly reliable pattern recognition systems, coming from the hybrid intelligent systems approach. Combination of several classifiers may result in improved performances [4], [5]. The necessity of combining multiple classifiers is arising from the main demand of increasing quality and reliability of the final models. There are different classification algorithms in almost all the current pattern recognition application areas, each one having certain degrees of success, but none of them being

as good as expected in applications. The combination technique we propose for the toxicity classification is a neuro-fuzzy gating of the implied classifiers, trained against the correct data. This approach allows multiple classifiers to work together.

For this task, the hybrid intelligent system NIKE was used, in order to automate the processes involved, from the data representation for toxicity measurements, to the prediction of toxicity for given new input. It also suggests how the fuzzy inference produced the result, when required [17], based on the effect measure method to combine the weights between the layers of the network in order to select the strongest input-output dependencies [6]. Consequently, for NIKE, we defined the *implicit knowledge* as the knowledge acquired by neural/neuro-fuzzy nets.

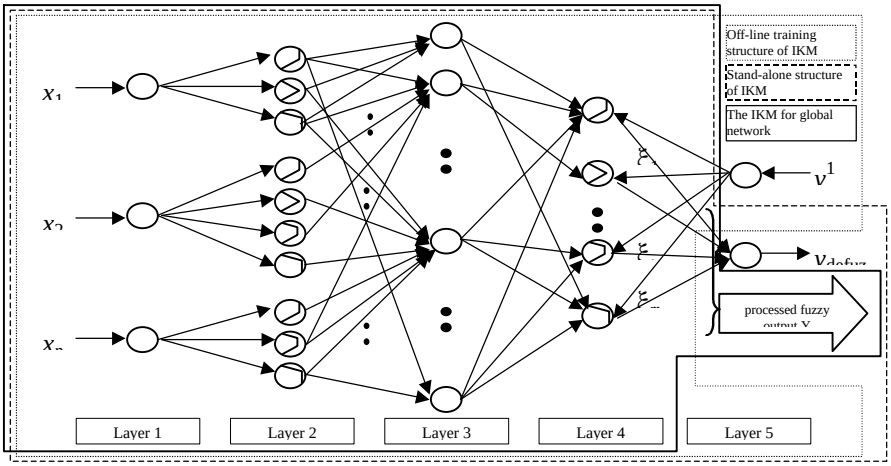


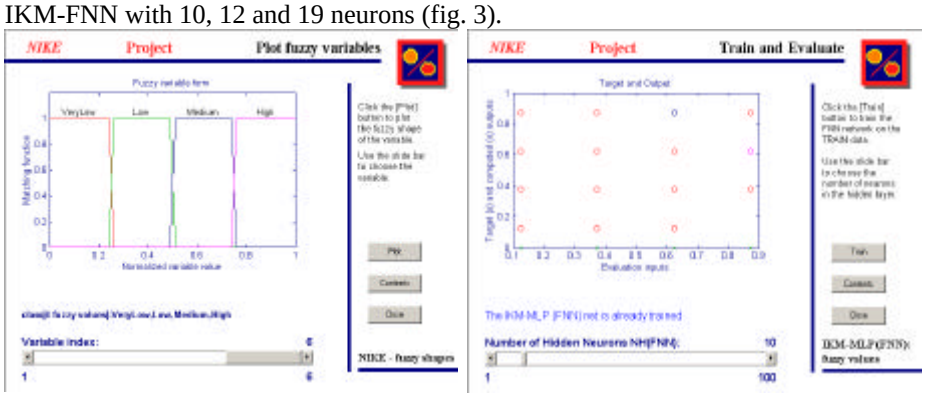
Fig. 1. Implicit Knowledge Module implemented as FNN2.

The IKM-FNN is implemented as a multilayered neural structure with an input layer, establishing the inputs to perform the membership degrees of the current values, a fully connected three-layered FNN2 [16], and a defuzzification layer [17] (fig.1). The weights of the connections between layer 1 and layer 2 are set to one. A linguistic variable X_i is described by m_i fuzzy sets, A_{ij}, having the degrees of membership performed by the functions $\mu_{ij}(x_i)$, j=1,2,...,m_i, i=1,2,..,p., (in our case, p=5, all m_i=4, on the classes of the prediction result of the classifiers, as inputs, and on the classes of the toxicity values, as the output y_{defuz}). The layers 1 and 5 are used in the fuzzification process in the training and prediction steps, and the layers 2-4 are organized as a feedforward network to represent the implicit rules through FNN training [15][19].

3.2 Results

Since NIKE modules process only data scaled into the interval [0..1], every class was represented by the centroid of each of the four classes in which the available domain was split: 0.135 (class 1), 0.375 (class 2), 0.625 (class 3), and 0.875 (class 4). The inputs and the output followed a trapezoidal (de)fuzzification (fig. 2): *VeryLow* (0-0.25), *Low* (0.25-0.5), *Medium* (0.5-0.75), *High* (0.75-1).

The neuro-fuzzy network was trained on a training set of 40 objects (70% of the entire set, as depicted in Table 2). The training set was used for the adjustment of the connections of the neural and neuro-fuzzy networks with backpropagation (traingdx) algorithm; *traingdx* is a network training function that updates weight and bias values according to gradient descent momentum and an adaptive learning rate. The neuro-fuzzy network was a multi-layered structure with the 5x4 above described fuzzy inputs and 4 fuzzy output neurons, the toxicity class linguistic variable (fig. 2.a). The number of hidden neurons parameterized the FNN. After different models (5 to 50 hidden units), a medium number of hidden units is desirable and had the same best results: IKM-FNN with 10, 12 and 19 neurons (fig. 3).



(a) (b)

Fig. 2. NIKE: (a)The fuzzy terms of the generic linguistic variable Class; (b) the FNN model.

Table 3. Performances of the classification algorithms computed.

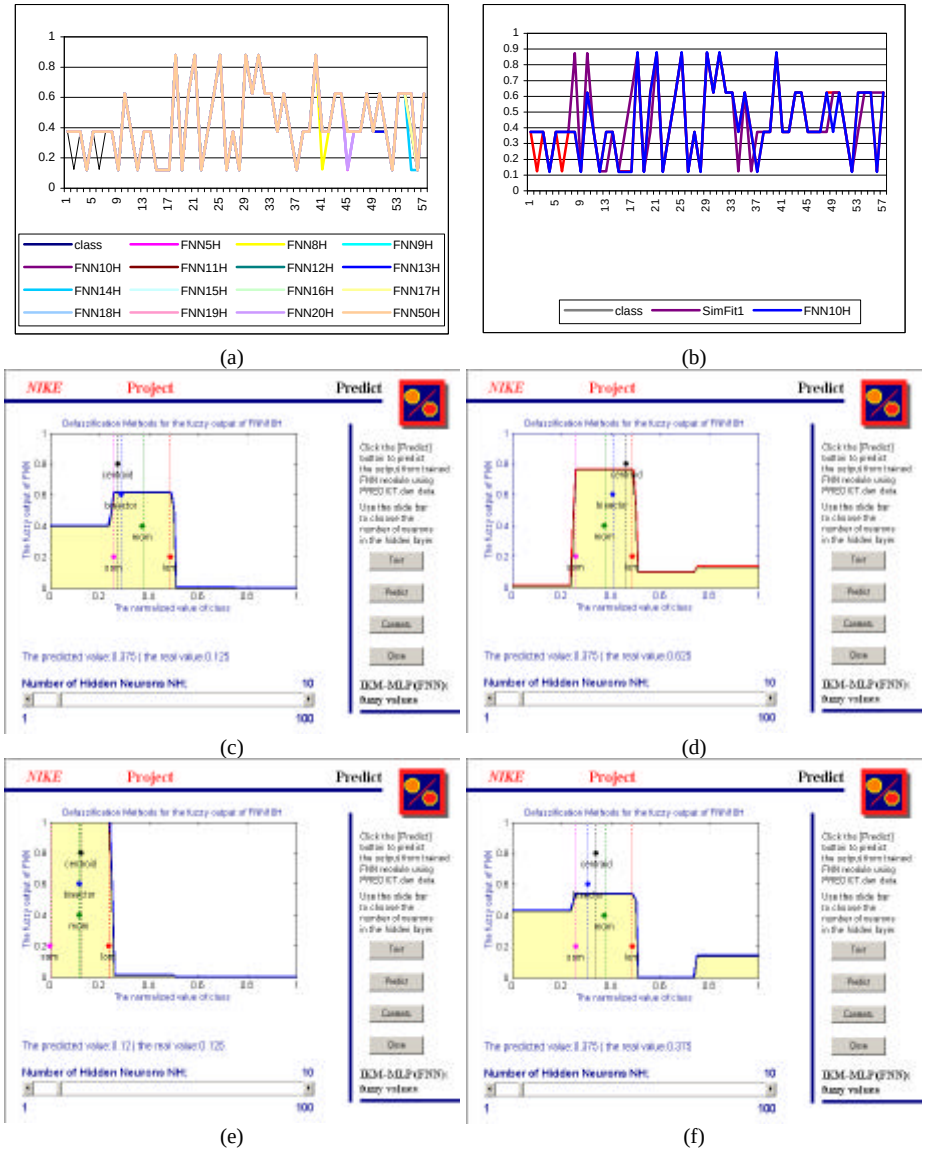
	NER% fitting	NER%validation	Descriptors
LDA	64.91	61.40	D1,D2, D3, D4
RDA	84.21	71.93	D1, D2, D3, D4, D6, D7, D8, D11, D12, D13
SIMCA	92.98	77.19	D1, D2, D3, D4, D5, D6, D7, D8, D10, D11, D12
KNN	-	61.40	D1, D12
CART	85.96	77.19	D1, D2, D3, D4, D5, D9

Table 4. Confusion matrix of the neuro-fuzzy combination of classifiers.

		Assigned Class				**N° of objects**
		1	**2**	**3**	**4**	
True Class	**1**	13	2			15
	2		20			20
	3		1	15		16
	4				6	6

Table 5. True class and class assigned by all the classifiers for each compound wrong predicted by the neuro-fuzzy combination of classifiers.

	True Class	**CART**	**LDA**	**KNN**	**SIMCA**	**RDA**	**FNN**
Chlorpyrifos	1	2	2	1	2	2	2
Profenofos	1	2	2	1	2	2	2
Fenitrothion	3	2	3	3	3	3	2



(a) (b)

(c) (d)

(e) (f)

Fig. 3. The results of training FNNs: (a) 3-5 errors, the best are FNN10H, FNN12H and FNN19H; (b) the chosen model, FNN10H, against the SIMCA results and the real ones; (c) the bad fuzzy inference prediction for 2 cases in class 1 (Chlorpyrifos and Profenofos); (d) the bad fuzzy inference prediction for the case in class 3 (Fenitrothion); two samples of good prediction for test cases: (e) a class 1 sample (Phorate); (f) a class 2 sample (Edinfenphos).

A momentum term of 0.95 was used (to prevent too many oscillations of the error function). The nets were trained up to 5000 epochs, giving an error about 0.015. The recognition error for the above models is 5.26% (table 4, 5, fig. 3).

The confusion matrix shows the ability in prediction of our approach. Looking of Table 3, we notice that the best performance was obtained by SIMCA, which could correctly classify almost 93% of the molecules. This encouraging result was obtained with whole data set involved in developing the model. If we take a look to the NER% validated with LOO, we can notice that we loss a lot of the reliability of the model when we predict the toxicity of an external object. Such a behavior proves the ability in modeling of these algorithms, but shows also their incapacity in generalization. The neuro-fuzzy approach seems to overcome this problem, succeeding in voting for the best opinion and underling all the considered classification algorithms (fig. 3).

3.3 Interpreting the results of the neuro-fuzzy combination of the classifiers

The most relevant fuzzy rules were extracted from the IKM-FNN structures using Effect Measure Method [6][13]. Finally, after deleting the contradictory rules, the next list of the most trusty fuzzy rules were considered for the chosen net IKM-FNN10H:

```
IF CarFit1 is:VeryLow THEN class is:High    (39.22%)
IF CarFit1 is:Low     THEN class is:High    (82.30%)
IF CarFit1 is:Medium  THEN class is:High    (48.74%)
IF CarFit1 is:High    THEN class is:High    (39.04%)
IF SimFit1 is:VeryLow THEN class is:Medium  (61.25%)
IF SimFit1 is:Low     THEN class is:Medium  (36.04%)
IF SimFit1 is:High    THEN class is:Medium  (43.72%)
IF RdaFit1 is:VeryLow THEN class is:Low     (75.65%)
IF RdaFit1 is:Low     THEN class is:Low     (100.00%)
IF RdaFit1 is:High    THEN class is:High    (76.39%)
```

Three types of fuzzy rules were obtained: some could be grouped by the same output, or by having the same fuzzy term in the premise and conclusion, and, finally, rules with mixed terms in premises and conclusion parts. From the first two groups of fuzzy rules (italics), we could conclude that, the opinion of the entry classifier is not important for the given output. More precisely, CART prediction for High values of toxicity (class 4) is better to not be taken in consideration.

```
IF (CarFit1 is:VeryLow) OR (CarFit1 is:Low) OR (CarFit1 is:Medium) OR
(CarFit1 is:High) THEN class is:High
```

Similarly, SIMCA outputs are not so important for predicting class 3 (Medium toxicity: the second group of fuzzy rules). From the second last group of rules, we could find which is the best classifier from the involved systems. In our case, in order to predict class 2 (Low toxicity) is better to consider the opinion coming from RDA. The same opinion is very important for predicting the class 4 (High toxicity) cases too.

Conclusions

Classification of the toxicity requires a high degree of experience from computational chemistry experts. Several approaches were described to generate suitable computer-

based classifiers for the considered patterns. We investigated five different classifiers and a neuro-fuzzy correlation of them, to organize and classify toxicity data sets. Our approach shown an improved behaviour as a combination of classifiers. Some results viewing fuzzy rules extraction, as well as the possibility to interpret particular inferences suggest that the Neuro-Fuzzy approach has the potential to significantly improve common classification methods for the use in toxicity characterization.

Acknowledgment

This work is partially funded by the E.U. under the contract HPRN-CT-1999-00015.

References

1. Benfenati, E., Pelagatti, S., Grasso, P., Gini, G.: COMET: the approach of a project in evaluating toxicity. In*:* Gini, G. C.; Katritzky, A. R. (eds.): Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools. AAAI 1999 Spring Symposium Series. AAAI Press, Menlo Park, CA (1999) 40-43
2. Benfenati, E., Piclin, N., Roncaglioni, A., Varì, M.R.: Factors Influencing Predictive Models For Toxicology. SAR and QSAR in environmental research, 12 (2001) 593-603.
3. Bishop, C.M.: Neural networks for pattern recognition. Clarendon Press, Oxford (1995)
4. Chen, K., Chi, H.: A method of combining multiple probabilistic classifiers through soft competition on different feature sets. Neurocomputing 20 (1998) 227-252
5. Duin, R.P.W., Tax, D.M.J.: Experiments with Classifier Combining Rules. Lecture Notes in Computer Science, 1857, Springer-Verlag, Berlin (2000) 16-29
6. Enbutsu, I., Baba, K., Hara, N.: Fuzzy Rule Extraction from a Multilayered Network, in Procs. of IJCNN'91, Seattle (1991) 461-465
7. Gini, G., Benfenati, E., Boley, D.: Clustering and Classification Techniques to Assess Aquatic Toxicity. Procs. of the Fourth Int'l Conf. KES2000, Brighton, UK, Vol. 1 (2000) 166-172
8. Gini, G., Giumelli, M., Benfenati, E., Lorenzini, P., Boger, Z.: Classification methods to predict activity of toxic compounds, V Seminar on Molecular Similarity, Girona, Spain (2001 forthcoming)
9. Gini, G., Lorenzini, M., Benfenati, E., Brambilla, R., Malvé, L.: Mixing a Symbolic and a Subsymbolic Expert to Improve Carcinogenicity Prediction of Aromatic Compounds. In: Kittler,J.,Roli,F.(eds.):*Multiple Classifier Systems*, Springler-Verlag, Berlin (2001)126-135.
10. Helma, C., Gottmann, E., Kramer, S.: Knowledge discovery and data mining in toxicology. Statistical methods in medical research, 9 (2000) 131-135
11. Ho, T., Hull, J., Srihari, S.: Decision combination in multiple classifier systems. IEEE Trans. Pattern Anal. Mach. Intell. 16/1 (1994) 66-75
12. Jacobs, R.A.: Methods for combining experts' probability assessments, Neur. Comp. 7/5(1995)867-888
13. Jagielska, I., Matthews, C., Whitfort, T.: An investigation into the application of ANN, FL, GA, and rough sets to automated knowledge acquisition for classification problems. Neurocomp,24(1999)37-54
14. Kosko, B.: Neural Networks and Fuzzy System. Prentice-Hall, Englewood Cliffs (1992)
15. Lin, C.T., George Lee, C.S.: Neural - Network Based Fuzzy Logic Control and Decision System. IEEE Transactions on Computers, 40/12 (1991) 1320-1336
16. Nauck, D., Kruse, R.: NEFCLASS-X: A Neuro-Fuzzy Tool to Build Readable Fuzzy Classifiers. BT Tech. J. 16/3 (1998) 180-192.
17. Neagu, C.-D., Avouris, N.M., Kalapanidas, E., Palade, V.: Neural and Neuro-fuzzy Integration in a Knowledge-based System for Air Quality Prediction. App Intell. J. (2001 accepted)
18. Palade, V., Neagu, C.-D., Patton, R.J.: Interpretation of Trained Neural Networks by Rule Extraction, Procs. of Int'l Conf. 7th Fuzzy Days in Dortmund (2001) 152-161.
19. Rumelhart, D.E., McClelland, J.L.: Parallel Distributed Processing, Explanations in the Microstructure of Cognition. MIT Press (1986)