Training through European Research Training Networks - Analysis of IMAGETOX

Giuseppina C. Gini
Dept of Electronics and Information
Politecnico di Milano
Piazza L. da Vinci 32, Milano.
Italy
gini@elet.polimi.it

Emilio Benfenati
Istituto di Ricerche Farmacologiche Mario Negri
Via Eritrea 62, Milano
Italy
benfenati@marionegri.it

Ciprian Daniel Neagu[1]
Dept of Electronics and Information
Politecnico di Milano
Piazza L. da Vinci 32, Milano.
Italy
neagu@fusberta.elet.polimi.it

***Abstract* – The European Training Networks are financed to help the cooperation and the exchange of young researchers through a network working on a research topic. Intelligent Modelling Algorithms for the General Evaluation of TOXicities (IMAGETOX) is a Research Training Networks, established in 2000 by EU under the Fifth Framework Programme. The training network exploits e-work and e-learning at researchers level. The organization and first experience of one of such network are illustrated. In particular we address here some issues about the concept and implementation of a Virtual Laboratory for Toxicology, a new concept deriving from the old tradition of Computer Chemistry.**

I. INTRODUCTION

The EU training network is an organization in which several centers collaborate and provide inputs to the training process in an integrated way. The advantages of this approach are particularly relevant when several domains are connected, because the best experiences on the individual domain are directly available from the relative center, in a way which is critically adapted, and so better usable, to be processed within the other domains. In this case, cross-fertilisation reaches higher results.

The concept of virtual and distance learning is extended to several reference centers. Typically distance learning refers to a single center, which provides the expertise to trainees. It is also possible to imagine a trainee which uses more than one distance learning center, but this does not mean that these centers are interconnected. The fundamental difference in the training network is that the member centers are strictly linked, and the activities are organised, offering a unique, advanced, integrated learning platform.

The concept of virtual laboratory is well established in chemistry, which is perhaps the science most used to reason on computer models, but is quite new in toxicology. In our case the virtual laboratory integrates inputs from different domains, within a single software product. Ecotoxicology provides the knowledge on mechanisms, rules, data, characterised by quality levels. Chemistry offers knowledge on chemical descriptors and physico-chemical properties. Computer science integrates these items. Additionally, it extracts new knowledge from the provided databases. The process at this point is not at all finished, because the centers with experience in toxicology and chemistry receive inputs from the proposed models, and elaborate the new results in an iterative way.

The paper is so organized: in section II we illustrate the scientific area for training, in Section III we give more details about the centers, in Section IV we explain the training needs in this area. In Section V we illustrate the web site developed, while in Section VI we describe our "virtual laboratory" based on NIKE, a neuro-fuzzy hybrid system.

II. THE RESEARCH THEME: QSAR AND ECOTOXICOLOGY

We are becoming increasingly aware of the need to understand and predict the consequences of chemicals to human health and the environment. The EU Council adopted Regulation (EEC) 793/93 on existing chemical substances. This Regulation consists of three main parts: data collection, priority setting and risk assessment. Given the large number of data gaps, it is the intention of the European Commission to use predictive methods for priority setting and, to a lesser extent, for risk assessment, within the framework of Council Regulation (EEC) 793/93.

The motivation of our project is to develop innovative predictive models, using advanced software tools, in order to obtain improved applicability of these systems. To this end, we develop the following objectives:

- To collect and organise information on different toxicological data bases, molecular descriptors, computational chemistry software, and statistical

[1] On leave from University Dunarea de Jos, Galati, Romania.

algorithms

- To propose some credible toxicity data sets to be used by the scientific community working on modelling and classification.
- To define the potential and limits of individual computational approaches
- To define the usefulness of different chemical descriptors.
- To verify the robustness of reported and new predictive models using independent sets of compounds not considered in the training set.
- To evaluate models that classify chemicals (i.e. toxic or non-toxic) as alternatives to methods that predict continuous values. This may be useful in light of the guidelines for (eco)toxicity in the European Union, which define toxicity classification, for the regulatory assessment of pesticides and drugs, and for the labeling of chemicals according to safety regulation.

In toxicity prediction there are many variables: the toxicological endpoint, the number of molecules in the data set, the homogeneity of the data set, the methods to describe the physico-chemical properties, the computational algorithm to produce the relationship, and the validation method are the principal ones.

Appropriate training has to be done keeping in mind the specific training in a given scientific field, but also the related fields. Modern research is more and more projected into complex fields, in which it is possible to gain major advantages by techniques developed for other aims. Such a cross-fertilization process will become a necessary tool to introduce innovative features. Young researchers have been trained in this direction.

In most Quantitative Structure Activity Relationships (QSARs) the uncertainty of the biological data has not been considered. However, it is well known that toxicological data have a given variability, due mainly to animal variability and to the use of different protocols. In some cases, also in authoritative data bases, toxicity data for a single compound, assessed with the same test, have been found to differ by up to three orders of magnitude. This project will consider this uncertainty, ensuring the reliability of the data, in order to avoid variability due to experimental protocols, but taking into account animal, and experimental, variability. This is a major difference between our approach and most of the previous models.

Most of the studies in the literature have used a single data set, a single approach to describe the chemical information (a limited number of molecular descriptors in most cases) and a single statistical algorithm. Indeed, there are so many and varied possibilities that an individual research group could not cope with all of them. Within this Research Training Network we are comparing different molecular descriptors and software used to calculate them (this can be performed more easily within a network).

It should be recognised that our aim is not to establish the best method for (eco)toxicity prediction. Such an outcome is unlikely. We believe that many methods can provide good predictive performance. The important objective is to assess the potential offered by each method. Successive steps will explore the feasibility of the integration of the different approaches. This is another innovative characteristic of our project. In this sense it is interesting to have not only the quantitative results of the predictions for comparison, but also the detailed background information. For instance, in terms of chemicals which are not predicted correctly, assessment can be made as to whether they are incorrectly predicted by all methods and as to whether the different models can be integrated in a more general system to take advantage of the potential of the different approaches.

Furthermore, we are evaluating methods to classify compounds according to (eco)toxicological properties in order to improve the applicability to real problems, such as ecotoxicological assessment.

The major concern over predictive models for toxicity is their applicability. It is well known that these models work well in many cases within the set of compounds and activities originally used, but their predictive ability is poor outside of the training set. In addition to the more common internal methods of validation, such as leave-one-out, bootstrapping, etc., we are constructing new data sets of compounds, to be used for external validation.

III. THE COLLECTIVE EXPERTISE

Below there is a description of each group of the network.

1: Istituto "Mario Negri", Milan, Italy , a scientific non-profit organisation for research and education. The institution, since it discovered the dioxin (2,3,7,8-TCDD) in Seveso soil in 1976, devotes a great deal of effort to environmental and toxicity studies.

2: School of Pharmacy and Chemistry, Liverpool John Moores University, UK, specialized in the prediction of toxicological activity from physico-chemical structure.

3: RITOX: Research Institute of Toxicology, Utrecht University, Faculty of Veterinary Sciences, working to understand the exposure and effects of individual chemicals and mixtures in the environment, and to develop predictive models, using a variety of modelling approaches and techniques .

4: Dipartimento di Elettronica e Informazione, Politecnico di Milano, Italy, active in Artificial Intelligence, in particular knowledge representation, planning, knowledge-based systems, constraint satisfaction systems; the application of the above.

5. NIC, National Institute of Chemistry, Ljubljana, Slovenia, with experience in chemometrics and computational chemistry.

6: UFZ-Umweltrforschungszentrum Leipzig-Halle GmbH, the only HGF centre where the work is devoted entirely to environmental sciences, also using QSAR.

7: Department of Chemistry, University of Tartu, active in quantum chemistry, the development and encoding of methods for the efficient use of statistical methods in

chemistry, and the development and introduction of molecular descriptors in structure – activity studies.

Collaboration is fundamental for the program. As stated above, the area of toxicity prediction involves expertise from toxicology, computational chemistry, organic chemistry, and computer sciences. No single group can cope with the many fields involved. Furthermore, we believe that no single model can solve the highly complex problem of toxicity prediction. Several approaches are possible, providing useful models to explain, at least, part of the problem. It is possible that the information obtained with the different models is not similar (i.e. different physico-chemical properties are not found to be important by different methods). This will be assessed in the project. The important point which stems from the above consideration is that it is very useful to evaluate the methods to combine the information originating from the different approaches. To achieve this collaboration is essential. There will be exchange of data sets and descriptors and software, between the different groups.

IV. TRAINING NEEDS

There are many reasons for Europe to support training in this area.

1. In the twenty-first century more work will have to be performed to appreciate, understand and predict the (eco)toxic effects of chemicals. The approach of predictive toxicology using advanced software is very novel and most of the universities in Europe have no adequate training in this field. The methodology is developing rapidly.
2. The characterisation of (eco)toxic properties is required by industry, who, by law, have to assess the toxicological hazards of their products and know the likely properties of future products, before starting synthesis and experimental studies
3. There is a requirement for this research from regulatory bodies and environmental agencies, who have to answer to questions from society about the safety of chemicals, and have to decide priority activities to prevent pollution.
4. This multi- and inter-disciplinary research will provide solutions to problems in the related fields of (eco)toxicology, computer science and computational chemistry. Young researchers will learn innovative approaches which will be highly valued in the workplace.

There is also a wider and more fundamental need for the EU to invest in training projects, such as this, which will provide the opportunity to increase the added value of a product, in this case a chemical. The Nobel Prize Winner Arno A. Penzias foresees that in the near future man will build molecules with desired properties. Europe has to invest in human potential in order to add new capabilities to products, to avoid scientists of European origin, such as Penzias, going abroad to prove the value of their ideas.

The training period is from 6 months to 3 years. The training programme is based upon environmental sciences, statistics, computer science, and computational chemistry. Further value in the programme arises from the combination of all these elements. Of course, the training will be focused on specific areas according to exact speciality of individual laboratories, but in all cases a common core knowledge will be provided to all young researchers.

In particular for the core training we have, periodically, short courses on the specific areas.[2]. Trainees actively participate to these meetings, and present their work. This allows them improving their communication skills as well as to share the knowledge obtained during the year. Young researchers are encouraged to attend and participate in conferences and congresses, on wider scientific areas, such as toxicology, computer sciences or environmental chemistry, to increase the generalisation of the specific work done during the project.

After 2 years the academic results are encouraging, with a high number of publications and presentations obtained, as illustrated in Fig. 1.

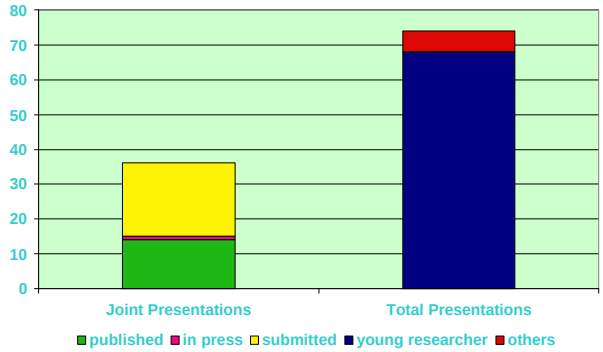

Fig 1. Number of presentations after 2 years

V. WEB BASED TECHNOLOGIES FOR KNOWLEDGE SHARING

Web-based promotion of the RTN IMAGETOX and dissemination of the RTN results is one of the targets proposed.

The web site has been developed by the third author as a collection of static pages representing the information at the first level of depth, ASCII flat databases, and hypermedia links to process Perl-implemented modules. The solution cgi (Common Gateway Interface) + Perl + ASCII flat databases was proposed in order to assure the simplest management and maintenance, without web administrator as well as web server solution dependences.

---

2 21/07/2001, a meeting within the V Girona Seminar, Spain. The meeting was attended by 11 participants from all the 7 groups.
24/01/2002-25/01/2002, IMAGETOX SCHOOL,. at National Institute of Chemistry, Ljubliana, Slovenia with 15 participants from all the 7 groups in the project and other scientific groups.

The idea could be extended to create, in the future, eventual mirror sites, if necessary. The Perl modules assure transparency in order to be checked as secure code by the web administrative staff. From the point of view of graphics design, the used files are small gif/jpg files, in order to assure small digital information to be downloaded. The web site[3] is best viewed with modern browsers (Netscape 4.0, MS Internet Explorer 5.0 and upper), on a resolution on minimum 800x600 pixels.

The other use of electronic tools is in the development of experiments. Data are processed in different partners sites, and exchanged, moreover the software tools developed at a site are freely transmitted.

VI. TOWARD A VIRTUAL LABORATORY FOR TOXICOLOGY PREDICTION

Until now, several papers have been published on the role that artificial intelligence (AI) could play in the problem of toxicity prediction and QSAR modelling [3]. In many cases learning from data is obtained using Neural Networks. In recent years, the neuro-fuzzy systems [2, 5] have drawn increasing research interest].

In this section we present neuro-fuzzy knowledge representation applied to toxicity prediction. The problem is modelled with NIKE (Neural explicit&Implicit Knowledge inference system), an hybrid intelligent system developed at Politecnico di Milano, and based on [7, 8]. The present implementation constitutes a virtual laboratory, based on a set of tools connected to the neuro-fuzzy architecture. Different modules can manage statistical models, QSARs, or trained networks. Moreover the system allows to extract symbolic knowledge from the trained neural networks, and to map many kinds of rules in neuro-fuzzy structures. These possibilities are the basis of a virtual laboratory for toxicology, where different predictive methods are usable, integrated, compared. Moreover the hybrid architecture allows combining the different models obtained to take the advantage of hybrid system to improve prediction [1].

The EU training network offers immediate possibility to partners to use and exploit the tool developed in one centre. The other centres cooperatively collaborated to develop their experiments. Below is an example of such a common experiment based on NIKE.

THE 'VIRTUAL' LABORATORY FIRST EXPERIMENT

The US Environmental Protection Agency [9] built a data set, starting from a revision of experimental data from literature, referred to acute toxicity 96 hours (LC_{50}), for fathead minnow (*Pimephales promelas*). The data set contains 568 organic compounds, commonly used in industrial processes.

A large number of descriptors was calculated by Istituto "Mario Negri". Molecular descriptors are a common way to represent knowledge about a chemical. The obtained descriptors were: *constitutional* (34); *geometrical* (14); *topological* (38); *electrostatic* (57); *quantum–chemicals* (6), and *hydrophobic* (7).Some of these descriptors do not add information, but increase the noise making more complex the result analysis. So descriptors selection was done with Principal Components Analysis (PCA) and Correlation Analysis techniques [6].

Three original QSAR equations were developed at Istituto "Mario Negri", from a different number of descriptors, using PLS (Partial Least Squares). The variables were normalized using auto scaling procedure: data were centred and standard deviation made equal to unit. Two models were obtained with five parameters (QSAR1, 2), a third model with two descriptors (QSAR3):

$$\log(1/LC_{50}) = 0.7919 + 0.09772*QM6 - 0.2045*C35 + 0.1276*G2 - 0.3509*pH9 - 0.3879*logP \quad (1)$$

$$\log(1/LC_{50}) = 0.8779 + 0.1385*QM6 - 0.06703*C35 - 0.02937*T6 - 0.06165*G12 - 0.6854*logP \quad (2)$$

$$\log(1/LC_{50}) = 0.8237 + 0.1711*QM6 - 0.7974*logP \quad (3)$$

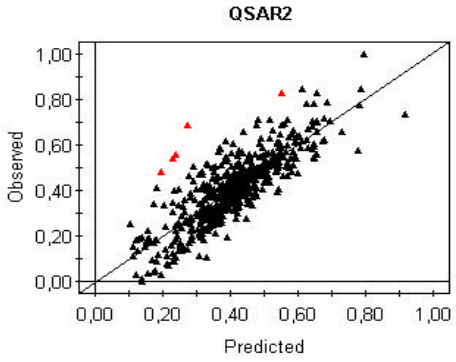

Figure 2. Observed versus predicted values (QSAR2).

Figure 2 reports the observed versus predicted values for LC_{50} relating to QSAR2. All models identified a restricted number of 5 outliers that do not satisfy the normal distribution required for residuals. Discharging these compounds, the performance increased reaching 70% o accuracy values (for an absolute error level of 0.1).

a. LEARNING FROM DATA WITH NIKE

Input and output values were fuzzified with respect to the 568 compounds descriptors values. The input contained 17 descriptors, the output toxicity expressed as $\log(1/LC_{50})$. For fuzzy processing, the membership functions were trapezoidal. The linguistic variables considered are characterized by the term sets:

$$D_i = \{Low, Med, High\}, i = 1..17 \quad (4)$$

$$\log(1/LC50) = \{VeryLow, Low, Medium, High, VeryHigh\} \quad (5)$$

Five levels of toxicity were defined for the normalized $\log(1/LC_{50})$: *VeryLow* (0-0.2), *Low* (0.2-0.4), *Medium* (0.4-0.6), *High* (0.6-0.8), and *VeryHigh* (0.8-1).

A pattern is defined as a vector of values of the inputs (descriptors) and of the output, toxicity. The training set

[3] http://www.airlab.elet.polimi.it/imagetox7

was used to train the neural and neuro-fuzzy networks with backpropagation. The whole set of patterns was randomly divided in two independent subsets, for training and testing, paying attention to conserve the distribution of the five fuzzy values of the output linguistic variable: 401 training cases and 167 testing cases (table 1).

Table 1. The distribution of testing + training sets.

Toxicity	VeryLow	Low	Medium	High	VeryHigh
Testing cases	15	66	72	12	2
Training cases	35	156	173	34	3
568 cases	50	222	245	46	5

b. NEURO-FUZZY STRUCTURES FOR TOXICITY REPRESENTATION

Implicit knowledge is the knowledge represented by neural/ neuro-fuzzy networks, created and adapted by a learning algorithm.

Explicit knowledge is the knowledge represented by neural networks, which are computationally identical to the I/O relations set, and are created by mapping the given fuzzy rules into hybrid neural networks.

i. The implicit knowledge representation

The module IKM-CNN (Implicit Knowledge Module-Crisp Neural Networks) models the data set as a multilayer perceptron (MLP) [2], and a procedure to extract an equivalent fuzzy-rules system is added, based on the interactive fuzzy operators [8]. The MLP model is also used to compare the performance of the neurosymbolic system with neuro-fuzzy and QSARs.

The module IKM-FNN (Implicit Knowledge Module-Fuzzy Neural Networks) is a multilayered neural structure with an input layer (to perform the membership degrees of the current values), a fully connected three-layered FNN2 [2], and a defuzzification layer. The weights of the connections between layer 1 and layer 2 are set to one. The 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., (on the descriptors and toxicity values). Since layers 1 and 5 are used in the fuzzification process, the layers 2-4 are organized as a feedforward network to represent the implicit rules through FNN training [2, 5].

Two steps were used to insert QSAR information in the implicit knowledge representation. This strategy follows an updated form of *concept support techniques* [7]. The pre-training phase is based on a data collection generated by a selected QSAR function. Then the model is trained with the original data set. The specific results are compared with the results coming from the normal training procedure, based on a random initialization of the weights of the neural networks. The neural and neuro-fuzzy networks resulted through QSAR insertion in a pre-training phase were retained for further combination of modules. The method is based on inserting the given QSAR2 predictions, and learning the training samples

The accuracy of prediction is a little better than QSARs. For the prediction accuracy, the error was calculated as the absolute value of the difference between the predicted and the actual value, for all observations. A second measure to compare the models is determining how many of the 568 observations were accurately predicted (absolute errorless than 0.1), relative to the number of cases (table 2).

Table 2**.** Comparison of the accuracy prediction for neural/ neuro-fuzzy structures: number of cases predicted with absolute error lower than 0.1

Toxicity	QSAR3	CNN	CNN*	FNN	FNN*
VeryLow	24	24	29	30	30
Low	189	196	203	205	201
Medium	210	200	198	212	219
High	30	28	26	27	31
VeryHigh	3	1	2	2	0

CNN*, FNN*: pre-trained/retrained neural networks.

ii. The explicit knowledge representation

The extended version of Modus Ponens

$$\frac{\text{IF } X_1 \text{ is } A_1 \wedge \ldots \wedge X_j \text{ is } A_j \text{ THEN Y is B} \qquad (X_1 \text{ is } A'_1) \wedge \ldots \wedge (X_j \text{ is } A'_j)}{\text{Y is B'}} \quad (6)$$

was used to infer results equivalent to the QSARs from the developed structures. This process is performed in four steps: 1) Matching (the compatibility between A' and A), 2) Aggregation (based on triangular norm), 3) Projection: the compatibility of (Y is B') with (Y is B) is obtained as an aggregation function; 4) Inverse-Matching and Defuzzification. Aggregation and projection are performed by generalized aggregative neurons, involving triangular norms or co-norms in Multi Purpose Neural Networks (MPNN) [2, 5].

There are two different types of relations mapped in MPNNs: empirical fuzzy rules about the descriptors, and QSARs.

The fuzzy rules are described as a discrete fuzzy rule-based system (DFRBS) in order to be mapped in MPNNs. The numerical weights corresponding to the connections between neurons are computed using Combine Rules First Method or Fire Each Rule Method [3].

To insert a single empirical fuzzy rule (Rule 1), we empirically developed a relation between toxicity and the important descriptors logP and QM6:

IF (*QM6* is *Low*) AND (*logP* is *High*) THEN $\log(1/LC_{50})$ is *Low* (7)

The implementation of the rule (7) is equivalent to a neuro-fuzzy network using MAPI neurons [7].

For inserting QSARs, the approximators given by implicit knowledge modules is replaced by the explicit knowledge module implementing a first-order Sugeno fuzzy model [10]. The output of MPNN is a single MAPI neuron, acting as an arithmetical device.

The integration of the developed structures follows the Fire-Each-Module Strategy (FEM), as proposed in general form in [4]. After off-line training of the implicit neuro-fuzzy module, the general output of the system is

composed as a T-conorm [8] of fuzzy outputs of each module: the four-layered IKM structure for global network and the EKM implemented using combine-rules-first or fire-each-rule method).

The system is equivalent to a set of fuzzy rules and the output is computed using firing rules first method [2]. Combination of the specific membership degrees provided by IKM and EKM is done here by the max fuzzy operator. In the hidden aggregative layer (AL), all the weights are set to one, and the neurons aggregate the computed membership degrees from the implicit and explicit modules.

In our case, the average of all the outputs is processed by AL+MAPI defuzzifier. The final neuron is a MAPI device, which computes the crisp value of the output. The developed modules are CNN*, FNN* (implicit knowledge), QSAR2 and QSAR3 (explicit knowledge). The final output based on FEM is the averaged output of the modules, as depicted in table3 and Figure 3.

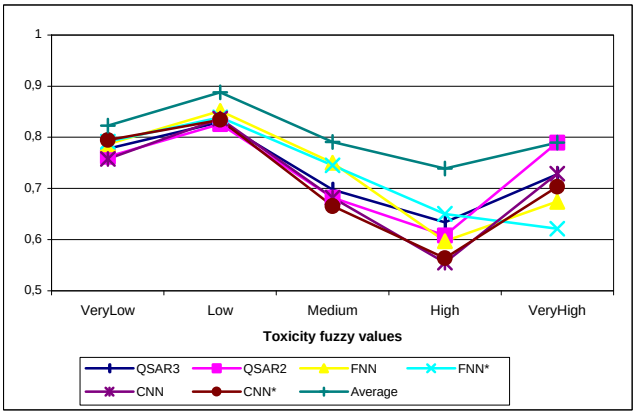

Figure 3. The accuracy of toxicity prediction, by classes (fuzzy values).

Table 3. Best results for accuracy of prediction (absolute error <0.1).

QSAR3	CNN*	FNN	FNN*	FEM
456	458	476	481	507
80.28%	80.63%	83.80%	84.68%	89.26%

VII. CONCLUSIONS

The flexibility and modularity of our approach allows coping with the complexity of the issues related with toxicity prediction. Other data sets have been tested, with different chemical descriptors and QSAR algorithms. NIKE proved to be able to manage them, mimic the QSAR models, and offering extra capabilities. Thus different research centres interacted with Politecnico di Milano for the use of NIKE. The peculiar issues of the different chemical and toxicological aspects were integrated with the hybrid system. In a second phase NIKE has been distributed within the network and this allowed multiplying the experience with the hybrid system. This shows a success history of a training network, as a result of an efficient organization based on a deep exchange of experiences and distance learning.

Moreover, Our study contributed to the understanding of the possibilities to represent knowledge about the toxicity of industrial organic compounds. The present approach represents an example of a hybrid system, combining artificial neural networks (ANN) and QSARs, on the basis of neuro-fuzzy modules implementation. The proposed neuro-fuzzy knowledge representation gives an encouraging alternative to the stochastic models; it proved to be able to learn from descriptors, and it is capable of representing knowledge acquired from human experts in order to improve the prediction results.

VIII. ACKNOWLEDGMENT

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

IX. REFERENCES

[1] E. Benfenati, P. Mazzatorta, D. Neagu, G. Gini, "Combining classifiers of pesticides toxicity through a neuro-fuzzy approach", *Proc. Multiple Classifiers Systems 2002*, Springer Verlag, 2002.

[2] R. Fuller, *Introduction to Neuro-Fuzzy Systems*, Advances in Soft Computing Series, Springer-Verlag: Berlin, 1999.

[3] G. Gini, A. Katritzky, "*Predictive Toxicology of Chemicals: Experiences and Impact of AI Tools*", AAAI Press, Menlo Park, Ca, 1999.

[4] R.A. Jacobs, M.I. Jordan, and A.G. Barto, 'Task decomposition through competition in a modular connectionist architecture: The what and where vision tasks', *Cognitive Science*, 15, 219–250, (1991).

[5] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*. Englewood Cliffs, NJ: Prentice-Hall, 1997.

[6] A.R. Katritzky, V.S. Lobanov, M. Karelson, *CODESSA Comprehensive Descriptors for Structural and Statistical Analysis. Reference Manual*, version 2.0, Gainesville, FL, 1994.

[7] C.-D. Neagu, M. Negoiþã, V. Palade, "Aspects of Integration of Explicit and Implicit Knowledge in Connectionist Expert Systems", *Procs 6th Int'l Conf on Neural Information Processing ICONIP'99* (T. Gedeon, et al, Eds.), Perth, 2, 759-764, (1999).

[8] C-D. Neagu, V. Palade, "An interactive fuzzy operator used in rule extraction from neural networks'" *Neural Network World Journal*, 10/4, 675-684, (2000).

[9] C.L. Russom, S.P. Bradbury, S.J. Broderius, D.E. Hammermeister, R.A. Drummond. "Predicting modes of toxic action from chemical structure: acute toxicity in the fathead minnow (*Pimephales promelas*)", *Environ. Toxicol. Chem.*, 16, (1997).

[10] M. Sugeno, G.T. Kang, "Structure identification of fuzzy model", *Fuzzy Sets and Systems*, 28, 15-33, 1988.