

# An Evolutionary Approach for Multi-Objective Feature Selection in ADMET Prediction

Axel J. Soto<sup>1,2,3</sup>, Rocío L. Cecchini<sup>1,3</sup>, Damián Palomba<sup>1</sup>, Gustavo E. Vazquez<sup>1</sup>,  
Ignacio Ponzoni<sup>1,2</sup>

<sup>1</sup> Laboratorio de Investigación y Desarrollo en Computación Científica (LIDECC),  
Departamento de Ciencias e Ingeniería de la Computación (DCIC),  
Universidad Nacional del Sur – Av. Alem 1253 – 8000 – Bahía Blanca – Argentina

<sup>2</sup> Planta Piloto de Ingeniería Química (PLAPIQUI),  
Universidad Nacional del Sur – CONICET  
Complejo CRIBABB – Camino La Carrindanga km.7 – CC 717 – Bahía Blanca – Argentina

<sup>3</sup> These authors contributed equally to this manuscript  
{saj, rlc, dp, gev, ip}@cs.uns.edu.ar

**Abstract.** This paper presents multi-objective evolutionary methods for determining the most relevant set of variables for predicting physicochemical properties. The multi-objective approach is useful to both minimize the cardinality of the subset as well as to maximize its predictive capacity. In this sense, rigorous experimentations were carried out in order to determine which multi-objective strategy is better for the feature selection task. Based on the results over a logP (octanol-water partition coefficient) data set, we may argue that the aggregative (non-Pareto) strategy constitutes a wise search strategy.

**Key words:** multi-objective, evolutionary algorithms, feature selection, logP.

## 1. Introduction

Historically, when a new drug had to be developed, a ‘serial’ process started where drug potency (activity) and selectivity were examined first [1]. Many of the candidate compounds fail at later stages due to ADMET (absorption, distribution, metabolism, excretion and toxicity) behavior in the body. ADMET properties are related to the way that a drug interacts with a large number of macromolecules and they correspond to the principal cause of failure in drug development [1].

Currently, the failure rate of a potential drug before reaching the market is still high. The main problem resides in the difficulty to know the rules that govern ADMET behavior in the human body. For these reasons, interest in Quantitative Structure-Property Relationships (QSPR) given by the scientific and industrial community has grown considerably in last decades. QSPR comprises the methods by which chemical structure parameters are quantitatively correlated with a well defined process, such as biological activity or any other experiment.

Computational prediction methods for ADMET properties are commonly named as *in silico* techniques. Although these methods are not pretended to replace *in vitro* experiments, at least in the short term, some computer methods have demonstrated to obtain as good accuracy as well-established experimental essays [2]. QSPR has evolved over a period of 30 years from simple regression models to different computational intelligence models that are now applied to a wide range of problems [2, 3]. Nevertheless, the accuracy of the ADMET property estimations remains as a challenging problem [4]. In this sense, machine learning methods are most preferred given the great amount of existing data and the little understanding of the pharmacokinetic rules of xenobiotics in the human body.

In this context, hydrophobicity is one of the most extensively modeled physicochemical properties since the difficulty of experimentally determine its value, and also because it is directly related with ADMET properties [3, 5]. This property is traditionally expressed in terms of the logarithm of the octanol-water partition coefficient (logP). One common way of expressing the structural composition of a molecule in a QSPR method is by the calculus of whole-molecular descriptors. Each descriptor defines an attribute of the entire molecule, and its value could be obtained by experimental measures or numerical methods [6]. One major dilemma when logP is intended to be modeled by QSPR is that, with the exclusion of a few common descriptors, there is no general agreement of which descriptors are relevant or influence the hydrophobic behavior of a compound. This is an important fact, because overfitting and chance correlation could occur as a result of using more descriptors than necessary [7, 8]. On the other hand, poor models come as a result when less descriptors than necessary are used. For this reason, both identifying the adequate number of selected descriptors as well as inferring most influential descriptors constitute an important challenge. In terms of Artificial Intelligence, this topic constitutes a particular case of the feature selection problem.

Thereby, this work presents a multi-objective evolutionary algorithm (MOEA) for inferring most influential descriptors for physicochemical properties. The correctness of the selection is assessed by an independent prediction model. Our technique is based in the application of MOEAs, where: different fitness functions and different evolutionary approaches are used. This paper is organized as follows: next section discusses related issues on feature selection. Section 3 expands the aforementioned idea by the introduction of the MOEAs proposed for descriptor selection. In Section 4, the design of the experiments is presented, followed by the results obtained for the logP prediction. Finally, in Section 5, main conclusions and future work are discussed.

## 2. Feature Selection

Feature selection (FS) is a process commonly used in machine learning for selecting from or reducing a set of variables used to describe a target variable in a dataset. Thereinafter, variables, features or descriptors will be used here without distinction.

Biological and chemical databases are in a continuous growth so reducing dimensionality in a learning task is critical; not only for reducing storage and processing requirements, but also it is important for improving understanding and visualization of relationships in data. Moreover, variables may be noise-dominated, redundant or ir-

relevant, therefore the learning process using raw variables would be detrimental for the prediction performance [9].

FS methods could be categorized in terms of whether they are employed in a supervised or in an unsupervised scenario [10, 11]. In the supervised scenario, the utility of the selection of features is usually assessed regarding a target variable. This assessment could be made in two ways: by determining the performance of a machine learning model at predicting using the reduced subset of variables (wrapper methods), or by using ‘proxy’ measures that consider how well the target variable is described by the distribution of the reduced feature values (filter methods) [9].

Supervised methods may be applied in two main ways, i.e. in terms of whether variables are individually or globally evaluated. When the first approach is applied the method ranks variables according to their individual predictive power. However, a variable that is useless by itself could be useful jointly with others variables [12]. Therefore, more powerful learning models are obtained, when the FS model selects subsets of variables that jointly have good predictive capacity. It is worth noting that the FS problem of selecting a subset from a set of  $d$  variables has a complexity of  $2^d$ , so brute force search becomes unfeasible even for moderate values of  $d$ . As a result, FS methods have to make use of heuristics or randomized search strategies.

Wrapper-based FS methods are usually conformed by two components: an objective function, which may be a learning (regression or classification) method and a searching function that selects variables to be evaluated by the objective function. The results of the objective function are used to guide the searching procedure in the selection of descriptors.

A wrapper method could also be analyzed as a multi-objective optimization: where the predictive capacity of the reduced set of variables is intended to be maximized, while maintaining the cardinality of the selected subset minimal. This approach has been previously investigated [13] and is essentially motivated by the fact that larger feature sets will usually result in a better prediction performance, but also the subset will be prone to overfitting and hence to have a poor generalization performance [11]. It is important to remark that for multi-objective wrappers, its objective component is integrated by two, or more, objective functions.

### 3. Multi-Objective Evolutionary Algorithms

Having in mind the minimization of the number of descriptors selected and the prediction error obtained with the selected subset, we implemented three Multi-Objective Evolutionary Algorithms (MOEAs) for searching the space of multiple feasible selections. The first method consists in an aggregative MOEA that combines error and number of descriptors in a single formula. The second and third methods use NSGA-II and SPEA2 Pareto-based strategies respectively [14]. Moreover, for guiding the search of the MOEAs we use three appropriate fitness functions based on decision trees,  $k$ -nearest neighbors (KNN) and a polynomial non-linear function. According to the previous classification, our proposed FS methods belong to wrapper methods because statistical or machine learning methods are used in the fitness functions for assessing the prediction capability of the selected subset.

### 3.1 MOEAs Characteristics

Binary strings are used to represent the individuals. Each string of length  $m$  stands for a feasible descriptor selection, where  $m$  is the number of considered descriptors. A nonzero value in the  $i^{\text{th}}$  bit position means that the  $i^{\text{th}}$  descriptor is selected.

The initial population is randomly generated. A one-point crossover is used for the recombination [15]. An independent bit mutation method is used to mutate the individuals (toggles each bit with probability  $p$ ). The selection scheme depends on the MOEA. For the aggregative strategy, we perform different experiments with typical selection methods and we concluded that tournament method is appropriate. Furthermore, this method is more preferred than others, because is particularly easy to implement and its time complexity is  $O(n)$  [15]. In the other two MOEAs, selection operator corresponds with the NSGA-II or the SPEA2 ones respectively. All three MOEAs included elitism, which protects the fittest individuals in any given generation, by moving them to the next generation.

### 3.2 Fitness Functions

Keeping in view that the MOEAs objectives are to minimize the number of selected descriptors and to determine the most relevant set of descriptors for predicting a physico-chemical property, we defined two objective functions. The first objective function  $F_1$ , calculates the number of ones in each individual, i.e. the number of selected descriptors. For the second goal, the function estimates the accuracy of a prediction method when a given set of descriptors is used. In particular, the function  $F_2$  for the  $k^{\text{th}}$  individual is presented in equation 1. This formula computes the mean square error of prediction (MSE).

$$F_2(\mathcal{P}_{Z_1^k}, Z_2^k) = \frac{1}{n_2} \sum_{(\bar{x}_i, y_i) \in Z_2^k} (y_i - \mathcal{P}_{Z_1^k}(\bar{x}_i))^2. \quad (1)$$

Where:

- $Z$  is a matrix that represents a compound dataset, where each row and column corresponds to a compound and a descriptor respectively. Last column of  $Z$  stores the experimental target values for each compound. This column vector is denoted as  $y$ .
- $\mathcal{P}_Z$  is a predictor method trained with the dataset  $Z$ .
- $Z_1$  and  $Z_2$  are compound databases, that are used as training and checking sets (see Section 4.2) respectively, with corresponding sizes  $n_1 \times m$  and  $n_2 \times m$ .
- $Z_i^k$  is a filtered dataset according to the descriptor selection encoded in the  $k^{\text{th}}$  individual. In other words,  $Z_i^k$  only contains those variables of  $Z_i$  whose values in the corresponding locations of the  $k^{\text{th}}$  individual's chromosome is 1.
- $\bar{x}_i$  is a vector that represents the values of the descriptors for the  $i^{\text{th}}$  compound of a given dataset.
- $y_i$  is the target value for the  $i^{\text{th}}$  compound of  $Z$ .

The first argument of the  $F_2$  is the predictor method applied to a given training set, while the second argument corresponds to a checking set, from where fitness value is calculated. As mentioned, three different predictors were used as  $\mathcal{P}_Z$ . The first one uses de-

cision trees for evaluating the predictive capacity of the selected descriptors encoded for a given individual. Specifically, the decision trees are used here as regression trees [16] without using any kind of pruning. The second is KNN regression as used in ref [17]. Both methods are local and usually applied for prediction or for FS purposes [18].

A non-linear regression model was also applied in this paper as first argument of the fitness function. A non-linear expression is established where their coefficients  $\beta_{i,j}$  are adjusted with a non-linear least-square fitting by the Gauss-Newton method [19]. The corresponding non-linear regression model formula is presented in Soto *et al.* [20]. Non-linear models are not generally applied given that they need the construction of a mathematical formula. Nevertheless, we proposed it as an alternative for neural networks, so that non-linear regressions could be carried out [21].

**Aggregative strategy.** This approach combines the objectives to derive a single criterion. As it is well known, these methods usually have good performance when tackling combinatorial optimization problems with convex Pareto fronts [22]. For the aggregative approach we propose the formula presented in equation 2. Here,  $\alpha$  is a relevance parameter for each objective, the first term of  $F_{AG}$  is intended exclusively to compute the prediction error and the second term reflects the ratio of selected descriptors scaled by  $F_2$ .

$$F_{AG} = \alpha F_2 + (1 - \alpha) F_2 \frac{F_1}{m}. \quad (2)$$

**Pareto strategies.** The NSGA-II and SPEA2 MOEAs were implemented within the PISA (Platform and Programming Language Independent Interface for Search Algorithm) framework [23].

#### 4. Methodology and Results

This work lies in the selection of a minimal-cardinality subset of descriptors which is intended to be used for predicting in a machine learning model. In particular, this proposal is applied for the prediction of the hydrophobicity of a chemical compound. Preceding works [20, 21] were proposed to put in evidence the advantages of using FS prior to the prediction of physicochemical properties. Both papers offer a fairly comparison among different objective functions embedded in the fitness function. Therein, the selection process aims to minimize the prediction error in a mono-objective way, so the number of considered descriptors has to be varied within each run to establish the most appropriated number of relevant variables.

In this work the selection process is carried out by means of the MOEAs described in the previous section and using either aggregative or Pareto strategies. The main benefit of this approach is that there is no need to fix the number of descriptors to be considered. Looking for the minimization of the number of features is expected to result in a better generalization [11, 24]. Moreover, this work is also aimed to obtain a rigorous comparison among searching functions in MOEAs for FS. Finally, an independent validation machine learning method is applied to assess whether a specific feature subset selection is appropriate for the prediction of the target variable.

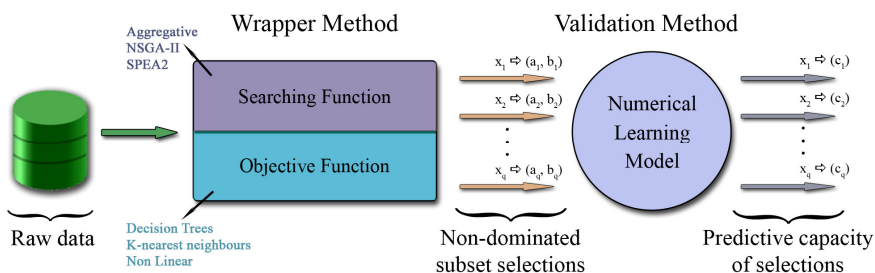
#### 4.1 Data Sets

We applied our MOEA FS on a data set [25] consisting of 440 organic compounds with 12 calculated molecular descriptors and the hydrophobicity as the output variable. We additionally incorporated 61 descriptors from the constitutional (40), functional groups (16), properties (2) and empirical (3) families. Previous to the use of the data, all descriptors were standardized. In contrast to [25], the cases were not exactly separated as specified therein, since given the data set size, the use of fixed data sets could lead to bias problems and therefore is not representative of the generalization capacity.

#### 4.2 Experimental design

As it was mentioned, the wrapper method is conformed by a searching and an objective function. The MOEAs were conceived for the searching functions. The MOEAs' parameters were intentionally kept equal, e.g. population size (45), crossover (0.8) and mutation probability ( $1/m$ ). In particular, the aggregative strategy has a weighting parameter  $\alpha$ , which was set into two alternative values (0.8 and 0.9) since we decide to give more importance to objective  $F_2$ . On the whole, with the NSGA-II and SPEA2 strategies, a total of 4 different MOEAs methods were applied.

In the same way, the three prediction techniques mentioned in section 3 were applied as the first argument of  $F_2$ . This makes 12 different combinations of wrapper methods, where for each one 5 replicas were obtained. Finally, we got the subset selections provided in the 60 non-dominated Pareto fronts, where each front may contain up to 45 individuals. Figure 1 shows the wrapper process for one independent run and the obtaining of a front of  $q$  non-dominated solutions. Principal Component Analysis (PCA) is applied prior to the application of the fitness function. This allows discarding linear redundant descriptors.



**Fig. 1.** Methodology outline of one independent run. All data are entered as input of the wrapper and  $q$  non-dominated solutions are obtained. A solution  $x_j$  has two objective function values  $(a_j, b_j)$ . A validation method is in charge of more precisely determining the predictive capacity of the selection. The wrapper is defined as a combination of a searching and an objective function.

It is worth to mention that each run of the MOEA uses a random split of cases in three sets: training, checking and support (to highlight the differences with the classical approach of training, validation and testing/hold-out). The first one is used by the fitness function for the creation of the regression model, while the second one is used for the application of the trained model in a different data set and the obtaining of the

fitness value. The support set is merged with the training and checking sets to be used in the validation method stage (Section 4.3).

### 4.3 Validation Method

The fitness functions applied to the MOEAs are not optimal methods for regression-based learning tasks. However, they are applied as objective functions here and elsewhere [18], since they provide a fast evaluation of the predictive capacity of the selection. Nevertheless, since wrapper methods are prone to overfitting, it is mandatory to finally assess the predictive capacity of the selection of the wrapper method with an independent validation method. This validation method is applied by training a learning model according to the selection of each non-dominated solution (Figure 1). Thereby, new data (the support set) are also added in order to not only use the over-represented set employed in the FS phase.

Neural networks (NNs) were chosen for the evaluation of the predictive capacity of the selected descriptors. NNs are probably one of the most widely used methods for QSPR regression [2, 3, 26]. To improve the stability of the NNs [27] we used an ensemble of neural networks consisting of 5 NN. The NN's architecture consists of a hidden layer with 5 units and the NNs were trained using Levenberg-Marquardt with bayesian regularization [27]. Input data is also normalized and transformed with PCA prior to the ANNE training.

### 4.4 Results

According to the experimental design of section 4.2, we obtained all the non-dominated solutions for each combination of searching and objective functions using 5 replicas within each combination. Each solution was evaluated using a 5-fold cross validation over the three data sets, thereby allowing a ranking of the most relevant features for predicting logP.

Given that our proposal consists of the comparison of the searching functions, we took the best selection obtained by each searching function (regardless of the objective function) and we also considered the  $\theta$  most frequently (MF) selected descriptors as a feasible subset selection, where  $\theta$  is the average number of selected descriptors. Additionally, we included the prediction results using all descriptors and a random selection of descriptors. Table 1 shows prediction results averaged on 20 replicas, and for these means we analyzed the significant differences based on three multiple comparison procedures, namely: DMS, Tukey and Bonferroni.

Tukey and Bonferroni procedures, which use a familywise error rate, and hence too conservative for the number of considered selections, found no significant differences, except for the best and worst selections. Looking at DMS results (comparisonwise error rate), we may conclude that any aggregative strategy is better than considering all descriptors. This latter claim can be stated with a Dunn Šidák familywise error rate of 0.0931.

It is worth highlighting that for all considered MOEAs selections, we found selections with lower subset cardinality that have better or comparable prediction capacity than predicting using all descriptors.

**Table 1.** Prediction errors, subset cardinality and multiple comparison procedure results. Significance level was set to 0.065

Method	Aggregative $\alpha=0.8$ (MF)	Aggregative $\alpha=0.9$ (MF)	Aggregative $\alpha=0.9$	Aggregative $\alpha=0.8$	Pareto NSGA-II	Pareto SPEA2	Pareto SPEA2 (MF)	All descriptors	Pareto NSGA-II (MF)	Random
Cardinality	34	35	31	35	29	33	19	73	21	38,75
MSE	0,1620	0,1713	0,1716	0,1716	0,1781	0,1820	0,1854	0,1890	0,1933	0,2426
DMS	a	ab	ab	ab	abc	bc	bc	c	c	d
Tukey/Bonf.	a	ab	ab	ab	ab	ab	ab	ab	b	c

Finally, we present an analysis for emphasizing that the selected descriptors (Table 2) are a reasonable selection in a theoretical physicochemical sense [28, 29]. It is known that the descriptors MW/AMW, nC/nCaR/nCs are relevant in the determination of logP given that there is a linear relationship among them and the lipophilicity. When nC was not selected, the descriptors nCaR and nCs were selected. The descriptor nCaR is an estimate of the descriptor nBnz, so possibly nCaR (in cooperation with MW/AMW) was prioritized over nBnz for the selection. In the same way, nCs is an estimation of the presence of hydrocarbon chains and of the number of saturated rings nR03, nR05-nR07, nR09-nR11 and yet again, almost certainly the first one (jointly with MW/AMW) was prioritized.

**Table 2.** List of most frequently descriptors according to the aggregative strategy ( $\alpha=0.8$ ). Descriptors with \* are scaled on carbon atoms.

Rank	Symbol	Definition	Freq.	Rank	Symbol	Definition	Freq.
1	nCs	number of total secondary C(sp3)	97%	38	nR11	number of 11-membered rings	46%
2	nN	number of Nitrogen atoms	89%	39	nCp	number of total primary C(sp3)	46%
3	VMC1	first-order valence molecular connectivity index	80%	40	nR=CX2	number of R=CX2	46%
4	nCaR	number of aromatic C(sp2)	77%	41	nSH	number of thiols	43%
5	nHAcc	number of acceptor atoms for H-bonds (N, O, F)	71%	42	nR=CHX	number of R=CHX	43%
6	PO	average polarizability	69%	43	MR	Ghose-Crippen molar refractivity	43%
7	Sp	sum of atomic polarizabilities [*]	69%	44	RBN	number of rotatable bonds	40%
8	nBT	number of bonds	69%	45	nROR	number of ethers (aliphatic)	40%
9	nO	number of Oxygen atoms	69%	46	nCXr	number of X on ring C(sp3)	40%
10	MW	molecular weight	66%	47	Hy	hydrophilic factor	40%
11	D_H	total dipole (hybridization)	63%	48	nAT	number of atoms	37%
12	EX	exchange energy (two-center term)	63%	49	nOht	number of tertiary alcohols	37%
13	Ms	mean electrotopological state	63%	50	Sv	sum of atomic van der Waals volumes [*]	34%
14	nAB	number of aromatic bonds	63%	51	nCIR	number of circuits	34%
15	nCL	number of Chlorine atoms	63%	52	RBF	rotatable bond fraction	34%
16	AMW	average molecular weight	60%	53	nDB	number of double bonds	34%
17	Mv	mean atomic van der Waals volume [*]	60%	54	nH	number of Hydrogen atoms	34%
18	nR06	number of 6-membered rings	60%	55	D_P	total dipole (point charge)	31%
19	D_S	total dipole (sum = point charge + hybridization)	57%	56	ELC	total electrostatic interact. (two-center term)	31%
20	Se	sum of atomic Sanderson electroneg. [*]	57%	57	nBR	number of Bromine atoms	31%
21	nBM	number of multiple bonds	57%	58	nX	number of halogen atoms	31%
22	nR03	number of 3-membered rings	57%	59	nR07	number of 7-membered rings	31%
23	VMC2	second-order valence mol. connectivity index	54%	60	Ui	unsaturation index	31%
24	nI	number of Iodine atoms	54%	61	nR10	number of 10-membered rings	29%
25	nOHs	number of secondary alcohols	54%	62	PSA	fragment-based polar surface area	29%
26	nRSR	number of sulfides	54%	63	E2	total two-center energy	26%
27	Ss	sum of Kier-Hall electrotopological states	51%	64	Mp	mean atomic polarizability [*]	26%
28	nTB	number of triple bonds	51%	65	nF	number of Fluorine atoms	26%
29	nC	number of Carbon atoms	51%	66	nCq	number of total quaternary C(sp3)	26%
30	nOHp	number of primary alcohol	51%	67	nCt	number of total tertiary C(sp3)	23%
31	IP	ionization potential	49%	68	nHDon	number of donor atoms for H-bonds (N & O)	23%
32	VMC4	fourth-order valence mol. connectivity index	49%	69	nR09	number of 9-membered rings	20%
33	Me	mean atomic Sanderson electroneg. [*]	49%	70	SCBO	sum of conventional bond orders (H-depl.)	17%
34	nBO	number of non-H bonds	49%	71	ARR	aromatic ratio	17%
35	nBnz	number of benzene-like rings	49%	72	nR05	number of 5-membered rings	11%
36	nSK	number of non-H atoms	46%	73	nS	number of Sulfur atoms	3%
37	nCIC	number of rings	46%				

The descriptors PO/SP are crucial since they are also linear with the dispersion interaction and thus with logP. The descriptor nCL is significant in terms of the fre-

quency of selection, and nCL is usually in closed relationship with lipophilicity; the descriptors nBR and nI were not often selected, but instead descriptor (nX) was either selected or probably its hydrophobicity effect was contemplated in the molecular weight descriptors. Descriptors D\_P, D\_H and D\_S are also selected and this fact is in agreement with the theoretical knowledge that with the increase of these descriptors, the logP value diminishes. It is worth to note that all searching functions follow a similar criterion in the selection of descriptors (tables of selections of other searching functions could be accessed from [30]).

## 5. Conclusions and discussion

This work presents a methodology to detect which descriptors are the most influential to the prediction of the molecule hydrophobicity. This paper proposes an improvement over our previous FS methodologies [20, 21], in the sense of it allows an automatic selection of the number of relevant features by the introduction of a multi-objective approach of the FS task.

Other key contribution lies in the rigorous comparison among different searching functions. This comparison let us to claim that the aggregative strategy is an advisable way to encourage FS for QSPR applications. The cause of this claim resides in that Pareto strategy also looks for the spacing of the solutions in the front. Thereby, this factor is not necessary in FS, since it is more interesting to obtain the selection that allows as best as possible an optimal prediction capacity.

Despite the fact that our proposal involves a brute force phase in evaluating all non dominated solutions with a neural network ensemble, it is important to mention that original complexity is much more significant (order of  $2^{73}$ ). This approach is not restricted to logP, thus, we plan to apply our approach to other physicochemical properties.

**Acknowledgments.** Authors acknowledge the ANPCyT from Argentina, for Grant N°11-1652, and SeCyT (UNS), for Grants PGI 24/N019, PGI 24/ZN15, PGI 24/ZN16.

## References

1. Selick, H.E., Beresford, A.P., Tarbit, M.H.: The emerging importance of predictive ADME simulation in drug discovery. *Drug Discov. Today*, 7, 109-116 (2002)
2. Agatonovic-Kustrin, S., Beresford, R.: Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *J. Pharmaceut. Biomed.* 22, 717-727 (2000)
3. Taskinen, J., Yliruusi, J.: Prediction of physicochemical properties based on neural network modelling. *Adv. Drug Deliver. Rev.* 55, 1163-1183 (2003)
4. Tetko, I.V., Bruneau, P., Mewes, H., Rohrer, D.C., Poda, G.I.: Can we estimate the accuracy of ADME-tox predictions? *Drug Discov. Today*, 11, 700-707 (2006)
5. Huuskonen, J.J., Livingstone, D.J., Tetko, I.V.: Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indices. *J. Chem. Inf. Comput. Sci.*, 40, 947-955 (2000)
6. Livingstone, D.J.: The Characterization of Chemical Structures Using Molecular Properties. A Survey. *J. Chem. Inf. Comput. Sci.*, 40, 195-209 (2000)

7. Tetko, I.V., Livingstone, D.J., Luik, A.I.: Neural network studies. 1. Comparison of overfitting and overtraining. *J. Chem. Inf. Comput. Sci.*, 35, 826-833 (1995)
8. Topliss, J.G., Edwards, R.P.: Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem.*, 22, 1238-1244 (1979)
9. Blum, A.L., Langley, P.: Selection of relevant features and examples in machine learning. *Artif. Intell.*, 97, 245-271 (1997)
10. Liu H., Yu L.: Towards integrating feature selection algorithms for classification and clustering. *IEEE T. Knowl. Data En.*, 17, 1-12 (2005)
11. Handl J., Knowles J.: Feature subset selection in unsupervised learning via multiobjective optimization. *IJCIR* 3, 217-238, 2006.
12. Guyon, I., Elisseeff, A.: An Introduction to Variable and Feature Selection. *JMLR*, 3, 1157-1182 (2003)
13. Deb, K., Raji Reddy, A.: Reliable classification of two-class cancer data using evolutionary algorithms. *BioSystems*, 72, 111-129 (2003)
14. Coello Coello, C.A., Lamont, G.B., Van Veldhuizen, D.: *Evolutionary Algorithms for Solving Multi-Objective Problems*. 2nd edn. Kluwer Academic Publishers, New York (2007)
15. Goldberg, D.E., Deb, K.: A comparative analysis of selection schemes used in genetic algorithms, in *Foundations of Genetic Algorithms*. Morgan Kaufmann, San Mateo, CA, pp. 69-93 (1991)
16. Breiman, L.: *Classification and Regression Trees*, Chapman & Hall, Boca Raton, (1993)
17. Li, L., Weinberg, C.R., Darden, T.A., Pedersen, L.G.: Gene selection for sample classification based on gene expression data: Study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics* 17, 1131-1142 (2002)
18. Trevino, V., Falciani, F.: GALGO: An R package for multivariate variable selection using genetic algorithms. *Bioinformatics* 22, 1154-1156 (2006)
19. Madsen, K., Nielsen, H.B., Tingleff, O.: *Methods for Non-Linear Least Squares Problems*. Technical University of Denmark, (2004)
20. Soto, A.J., Cecchini, R.L., Vazquez, G.E., Ponzoni, I.: A Wrapper-based Feature Selection Method for ADMET Prediction using Evolutionary Computing. In: Marchiori E., Moore J.H. (eds) *EvoBio 2008*. LNCS, vol. 4973, pp. 188-199. Springer, Heidelberg (2008)
21. Soto, A.J., Cecchini, R.L., Vazquez, G.E., Ponzoni, I.: An Evolutionary Approach for Feature Selection applied to ADMET Prediction. *Ibero Amer. J. Artif. Intell.* 12, 55-63 (2008)
22. Coello Coello, C.A.: Evolutionary Multiobjective Optimization: A Historical View of the Field, *IEEE Computational Intelligence Magazine* 1, 28-36 (2006)
23. Bleuler, S., Laumanns, M., Thiele, L., Zitzler, E.: PISA A Platform and Programming Language Independent Interface for Search Algorithms in Evolutionary Multi-Criterion Optimization. Fonseca, C.M., Fleming, P.J., Zitzler, E., Deb, K., Thiele L. (eds) *EMO 2003*, LNCS, vol. 2632, pp. 494-508. Springer, Heidelberg (2003)
24. Schmidhuber, J.: Discovering neural nets with low kolmogorov complexity and high generalization capability. *Neural Networks*, 10, 857-873 (1997)
25. Yaffe, D., Cohen, Y., Espinosa, G., Arenas, A., Giralt, F.: Fuzzy ARTMAP and backpropagation neural networks based quantitative structure - property relationships (QSPRs) for octanol: Water partition coefficient of organic compounds. *J. Chem. Inf. Comp. Sci.* 42, 162-183 (2002)
26. Winkler, D.A.: Neural networks in ADME and toxicity prediction. *Drug. Future* 29, 1043-1057 (2004).
27. Bishop, C.M.: *Pattern Recognition and Machine Learning.*, Springer Science + Business Media, New York (2006)
28. Morrison, R.T., Boyd, R.N.: *Organic Chemistry*. Prentice Hall, New Delhi (1992)
29. Chang, R.: *Chemistry*. McGraw Hill, New York (1998)
30. Laboratorio de Inteligencia y Desarrollo en Computación Científica (LIDeCC), Universidad Nacional del Sur. [http://lidecc.uns.edu.ar/downloads/CLEI08/Ranking\\_of\\_descriptors.xls](http://lidecc.uns.edu.ar/downloads/CLEI08/Ranking_of_descriptors.xls)