Prediction and Discrimination of Pharmacological Activity by Using Committees of Artificial Neural Networks

María José Castro Bleda[1] Wladimiro Díaz Villanueva[2] Juan Lucas Domínguez Rubio[1] Salvador España Boquera[1]

[1]Dep. de Sistemas Informáticos y Computación — Universidad Politécnica de Valencia
Camino de Vera s/n, 46071 Valencia, Spain

[2]Dep. de Informática — Universitat de València
Dr. Moliner 50, 46100 Burjassot (Valencia), Spain

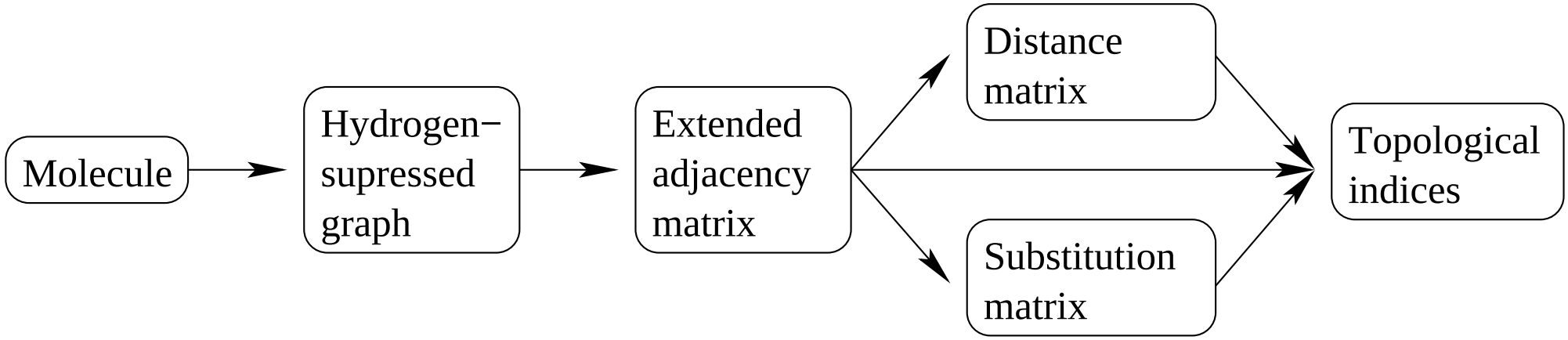
mcastro@dsic.upv.es wladimiro.diaz@uv.es jldoru@hotmail.com sespana@dsic.upv.es

ABSTRACT

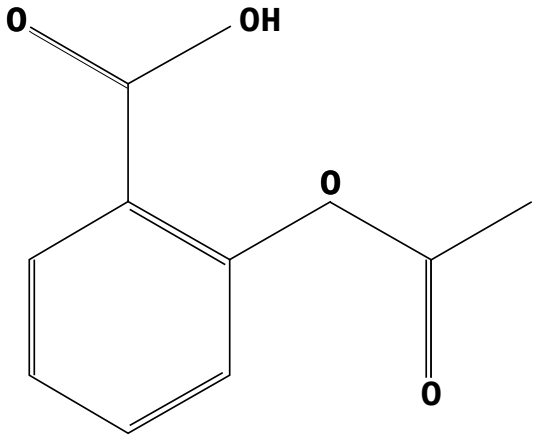
The design of new medical drugs is a very complex process in which combinatorial chemistry techniques are used. For this reason, it is very useful to have tools to predict and to discriminate the pharmacological activity of a given molecular compound so that the laboratory experiments can be directed to those molecule groups in which there is a high probability of finding new compounds with the desired properties.

A suitable set of topological indices that describe the molecular structure is used in this work. Two discrimination problems and two prediction problems are studied, using committees of multilayer perceptrons to discriminate/predict. A large amount of different configurations are tested, yielding very good performances.

REPRESENTATION OF THE MOLECULES



Process to obtain the topological indices used to represent the molecules



Molecular structure of (*aspirin*)

$\{9, 0, 4, 0, 0, 0, 0, 8, 5, 0, 4, 5, 4, 0,$
$4, 2, 0, 0, 0, 0, 0, 0, 0, 0, 5, 10, 8, 11, 7, 7, 0, 0, 0, 0, 0, 0, 0, 0, 0,$
$0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 13, 17, 16, 15, 11, 6, 0, 0, 0, 0, 0, 0\}$

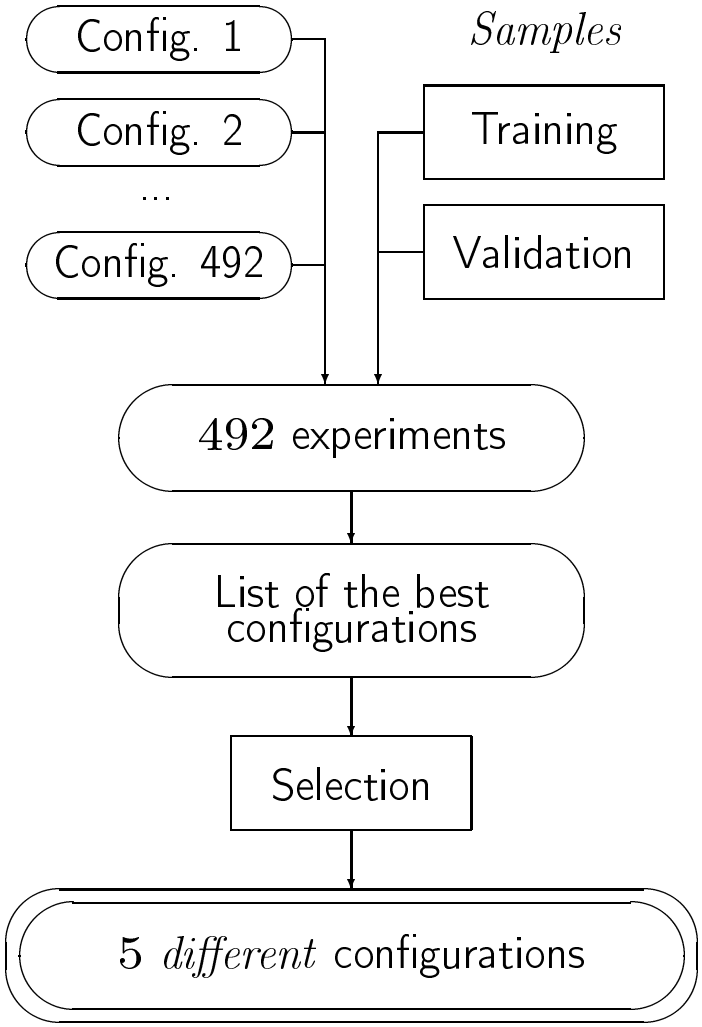
Set of topological indices (62 indices)

MLPs FOR PREDICTION AND DISCRIMINATION

DATA

C Classification problem: 2 classes
+ Positive samples (active +1)
− Negative samples (non-active -1)
P Prediction problem (quantitative)

		Number of samples		
Problem	*Type*	+	−	*Total*
Analgesic	C	172	813	985
Antidiabetic	C	180	163	343
MIC	P			111
Solubility	P			92



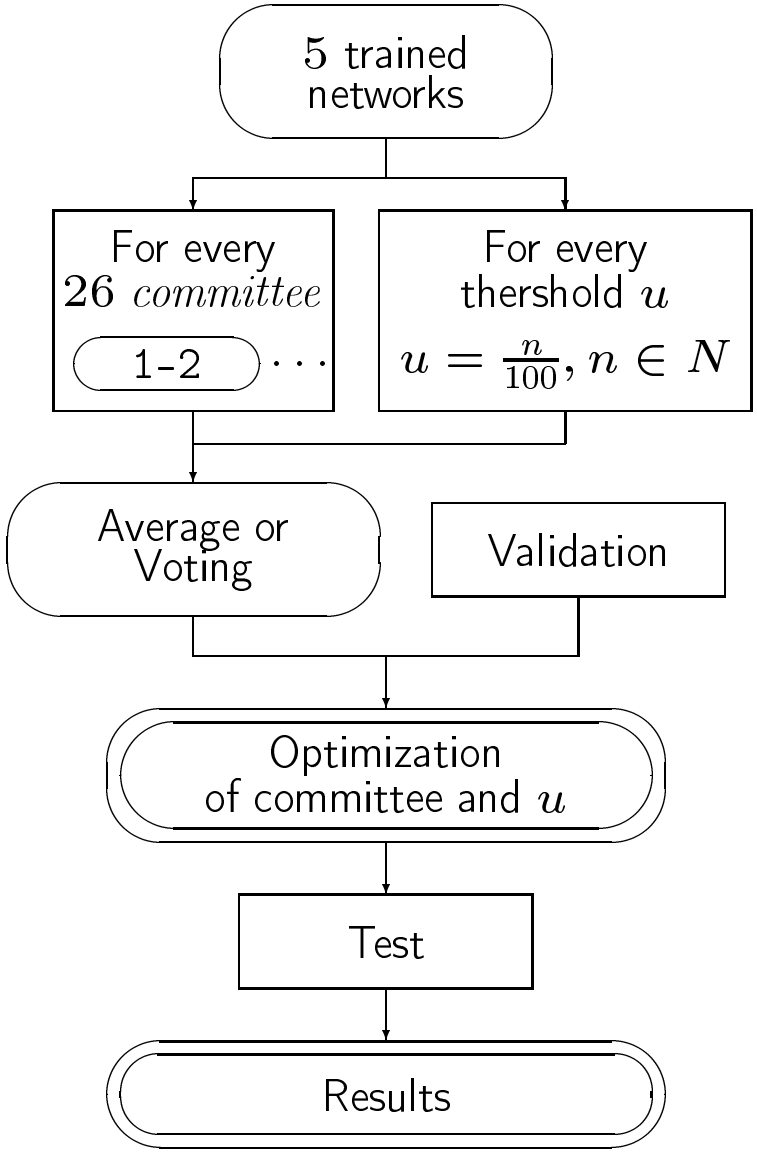
BEST INDIVIDUAL MLPs

Test results with a confidence interval of 95% (c.v. is 'central value').

Problem	*MLP*	*Test results Success rate (%)*
Analgesic	1	[81.31..89.94]
		c.v. 86.18
	2	[81.76..90.29]
	3	[82.21..90.63]
	4	[80.41..89.24]
	5	[79.96..88.89]
Antidiabetic	1	[87.10..97.49]
		c.v. 94.19
	2	[84.14..96.00]
	3	[88.64..98.18]
	4	[85.60..96.76]
	5	[85.60..96.76]

Problem	*MLP*	*Test results MSE*
MIC	1	0.0281869
	2	0.0281874
	3	0.0284059
	4	0.0284111
	5	0.0287237
Solubility	1	0.0103847
	2	0.0103955
	3	0.0105003
	4	0.0107689
	5	0.0111014

LOOKING FOR THE BEST COMMITTEE



BEST RESULTS

Prob.	*Success rate (%) or MSE*	*Prediction-classification*
Ana.	$86.99\,\%$	Committee composed by 3 nets with voting and threshold $u = 0.76$
Ant.	$94.19\,\%$	Individual MLP of topology 62–4–4–1 trained with *Backpropagation Momentum*.
MIC	0.45907	Individual MLP of topology 52–64–1 trained with *Standard Backpropagation*.
Sol.	1.74798	Individual MLP of topology 52–32–1 trained with *Standard Backpropagation*.

CONCLUSIONS

- ▷ In this work, the viability of the use of artificial neural networks for structure-activity discrimination and prediction has been shown. Two discrimination problems and two prediction problems were studied using the structural representation of the molecules.

- ▷ The high correlation of the neural network outputs does not allow to outperform the results by using committees of networks. Anyway, using thresholds can be of interest in some cases.