SciForum
MOL2NET

Artificial Neural Networks and Multilinear Least Squares to Model Physicochemical Properties of Organic Solvents

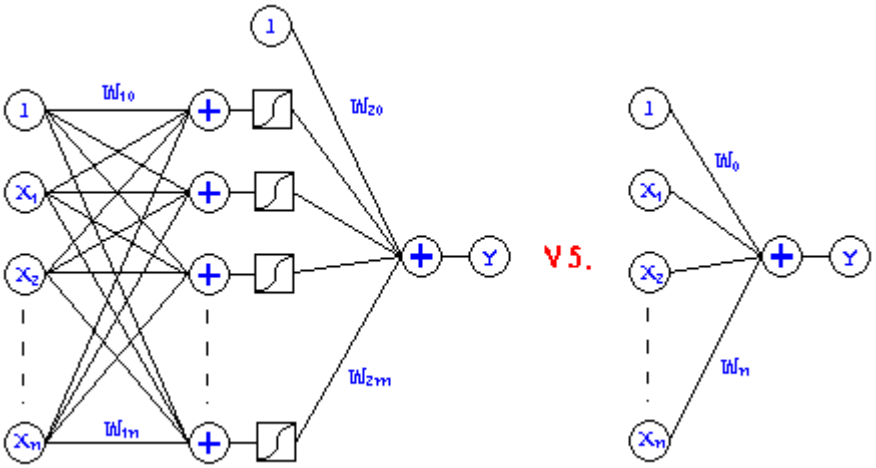
Jesus Vicente de Julián-Ortiz [1,2,*], Lionello Pogliani [1,3] and Emili Besalú [4]

[1] Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, 46100 Burjassot, València, Spain; E-Mails: jejuor@uv.es (J.V.d.J.O.); liopo@uv.es (L.P.);
[2] ProtoQSAR SL, Parc Científic, 46980 Paterna, València, Spain;
[3] MOLware SL, c/Burriana 36-3, 46005, València, Spain;
[2] Institut de Química Computacional i Catàlisi (IQCC) and Departament de Química, Universitat de Girona, Catalonia, Spain; E-Mail: emili.besalu@udg.edu;
* Author to whom correspondence should be addressed; Tel.: +34-963-543-279; Fax: +34-963-544-892.

Abstract: *The mean molecular connectivity indices (MMCI) proposed and used in previous studies are used here in conjunction with the well-known molecular connectivity indices (MCI) to remodel six properties of organic solvents. The MMCI and MCI descriptors of the Multilinear relationships for the six properties, obtained with the multilinear least - squares (MLS) procedure, were used to perform the artificial neural network (ANN) computations. The aim is to detect advantages and underline the limits of the ANN approach that, even if it improved the model, it is somewhat 'fuzzy' concerning the stability of the modeling. The MLS procedure replicates the obtained results as long as one wishes, a characteristic not shared by the ANN methodology, which, if on one side increases the quality of a description on the other increases also its overfitting. The present study reveals also how ANN methods prefer MCI relatively to MMCI descriptors. Four different types of ANN computations show that (i) MMCI descriptors are preferred with properties with poor number of points. MLS (ii) is to be preferred over ANN statistical results, with some exceptions, when the number of ANN weights is similar to the number of correlation coefficients of MLS. Furthermore, in (iii) some cases MLS modeling quality is quite similar to the modeling quality of ANN computations.*

Keywords: *Artificial Neural Networks, Multilinear Least Squares, physicochemical properties modeling, QSPR, molecular connectivity indices, mean molecular connectivity indices, boiling point, density, flash point, viscosity, surface tension, elutropic value.*

Graphical Abstract:



Introduction:

Recently [1] the mean molecular connectivity indices, MMCI, were introduced and used to model eleven properties of organic solvents. The multlinear least squares, MLS, used to derive the quantitative structure-property relationships (QSPR) showed that three out of six properties, the refractive index, *RI*, the flash points, *FP*, and the *UV* cutoff values, were modeled with the mean molecular connectivity indices, MMCI, while the remaining properties were modeled with the well-known molecular connectivity indices, MCI. The MMCI indices are centered on the basic concepts of the delta, valence delta, *I*- and *S*-indices that go back to the origins of the molecular connectivity theory. [2-7] Results from two other recent studies that used semiempirical sets of descriptors [8, 9] showed that the artificial neural network (ANN) procedure with variable number of hidden neurons, chosen by the software, normally improves the quality of a QSPR obtained by the aid of the MLS methodology. Nevertheless, this improvement is somewhat artificial as the ANN computations for the eleven properties employed a number of weights, due to the presence of more than one hidden neuron, much greater than the number correlation coefficients in the MLS procedure.

The present work aims are to pin down the real advantages but also the drawbacks of the ANN methodology applying it to the model of six properties of Ref.1 with only either MCI or MMCI used as descriptors. Four different types of ANN computations are here performed to detect the level of the achieved improvement, if any, (*a*) with one hidden neuron, (*b*) with a pre-fixed number of hidden neurons, (*c*) with a variable number of hidden neurons chosen by the software, and (*d*) with a minor number of descriptors, for the one hidden neuron case. This last case was tried to render the number of ANN weights equal to the number of correlation coefficients of the MLS case. Furthermore, it was monitored if ANN computations prefer either MCIs or MMCIs for modeling purposes. The descriptors for the six properties are those of ref. 1 but whenever a property was not satisfactorily

modeled by the given MCI (or MMCI) the second or third best MCI (or MMCI) were chosen.

Materials and Methods:

Table 1 shows the molecular connectivity χ indices, the molecular pseudoconnectivity ψ indices (*pseudo-MCI*), the dual connectivity and pseudoconnectivity indices (*Dual MCI, pseudo-MCI*) used in this study. Three new indices are also defined: $\Delta = \Sigma_{EA} n_{EA}$, $\Sigma = \Sigma_{EA} <S_{EA}>$, and $T_{\Sigma/M} = \Sigma^3/M^{1.7}$ (M = molar mass); Δ encodes the number of electronegative atoms (n_{EA}), Σ encodes the sum of the *S*-State index for the electronegative atoms, *N*, *O*, *F*, *Cl*, *Br* ($<S_{EA}>$ is the average value for a specific type of atom).

Table 2 shows the definitions of the MMCI (the first M stands for 'mean').

In Tables 1, and 2 i = 1-N denotes the atoms of a molecule, ij denotes directly σ-bonded atoms, and in Table 2, $p = N$. The Lehmer mean, ${}^{L}M$, for p = 2, equals the symmetrical mean, ${}^{S}M$. Replacing in Table 1 δ with the valence delta, δ^{v}, allows to obtain the corresponding valence MCI, $\{D^{v}, {}^{0}\chi^{v}, {}^{1}\chi^{v}, \chi^{v}_{t}, {}^{0}\chi_{d}{}^{v}, {}^{1}\chi_{d}{}^{v}, {}^{1}\chi_{s}{}^{v}\}$, while replacing the Intrinsic-*I*-State with the Electrotopological *S*-State index the corresponding pseudoconnectivity electro-topological indices are obtained, $\{{}^{S}\psi_{E}, {}^{0}\psi_{E}, {}^{1}\psi_{E}, {}^{T}\psi_{E}, {}^{0}\psi_{Ed}, {}^{1}\psi_{Ed}, {}^{1}\psi_{Es}\}$. [3-9] Replacing in Table 2 δ, with δ^{v}, *I* and *S* three other subsets of MMCI are obtained: the valence, $\{{}^{A}M^{v}, {}^{G}M^{v}, {}^{H}M^{v}, {}^{R}M^{v}, {}^{S}M^{v}, {}^{U}M^{v}, {}^{Ho}M^{v}, {}^{L}M^{v}, {}^{St}M^{v}\}$, the *I-State*, $\{{}^{A}M_{I}, {}^{G}M_{I}, {}^{H}M_{I}, {}^{R}M_{I}, {}^{S}M_{I}, {}^{U}M_{I}, {}^{Ho}M_{I}, {}^{L}M_{I}, {}^{St}M_{I}\}$, and the *E*-State $\{{}^{A}M_{E}, {}^{G}M_{E}, {}^{H}M_{E}, {}^{R}M_{E}, {}^{S}M_{E}, {}^{U}M_{E}, {}^{Ho}M_{E}, {}^{L}M_{E}, {}^{St}M_{E}\}$ MMCI, respectively. Because some *S* values can be negative (highly electropositive atoms) to avoid imaginary *S*-State MMCI values, a rescaling of the *S* value is undertaken as it is explained in ref. 7. Summing up, we have thirty-one MCI and thirty-six MMCI. Every index was obtained with a visual basic home - made program that runs on a normal PC that uses both adjacency and distance matrices [6].

The multilinear least squares procedure of Statistica 8 was used to find the best MCI and MMCI set of descriptors for the training set of Table 3, which is then used to evaluate the left-out compounds [those with (°) in Table 2, ~ 30% of all compounds, 25% for *El*]. The overall quality of each model was obtained with the Excel spreadsheet by plotting the observed versus the calculated property for the training and for the training plus evaluated points. The choice for the number of indices of a relationship has been done having in mind the *Topliss-Costello* rule: [10] the ratio of data points to the number of variables should be higher or equal to five, and should provide a correlation coefficient factor $r > 0.84$, i.e., $r^2 > 0.70$. The source for the properties of organic solvents listed in Table 3 is in ref. 7.

ANN methods perform regression and data validation, and carry out both tasks in a non-parametric way that makes no assumption regarding the relationship between y and x, where $y = f(x)$. This means that the function

Property = *f*(*indices*) is not known *a priori*. This non-parametric model is a kind of *black box* that tries to discover the mathematical function that can approximate the relationship between the *indices* and the *property* well enough. It uses highly flexible transfer functions with adaptable parameters that can model a wide spectrum of functional relationships.[11] The activation functions for both hidden and output nodes used in Statistica 8 are: identity (*i*), logistic sigmoid (*l*), hyperbolic tangent (*t*), sine (*s*), and exponential (*e*).

ANN results were obtained with the built-in utility of Statistica 8, the multilayer perceptron neural network (MLP). The used network has three-layer feedforward architecture with unidirectional full connections between successive layers and with error backpropagation (or backprop). The three layers are: *input units* → *hidden units* → *output units* (units are also known as neurons or nodes), and they correspond to: *variables* → *hidden units* → *P*, where the only output unit is the targeted property, *P*. In present study the number of variables corresponds to the number of MCI or MMCI descriptors. Each neuron (or node) in a particular layer is connected to every neuron in the next layer. The connections between neurons are practically the weights that determine the values assigned to the nodes. There exist additional weights assigned to the bias values that act as node value offsets, i.e., the weights of the connections: *input bias* → *hidden neuron*, and *hidden bias* → *P*. The number of weights is thus: [(*No. input nodes* + *2*)·(*No. hidden nodes*) + *1*]. The weights adjusted by the training process are initially random and are passed to all nodes of the following layer. The training process is iterative and each iteration is called an epoch. The weights are slightly varied in each epoch to minimize the sum-of-squares error function: $SOS = \sum_{i=1\text{-}N} (P_{iclc} - P_i)^2$, where P_{iclc} (*clc* = calculated) is the i^{th} predicted value (network outputs) of the property, and P_i is the target value. This function is the sum of differences between the prediction outputs and the target defined over the entire training set of points (compounds) *N*. Statistica 8 allows setting the number of networks to train and retain (*Ntr*/*Nre*). Here, two sets of values are chosen: *Ntr/Nre* = 10^3/200 and *Ntr/Nre* = 10^5/200. In the corresponding tables only *Ntr* is shown as *Nre* is constant. The ANN network of Statistica 8 optimized with the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm ensures a fast convergence rate.[12]

Statistica 8 initially sets by defect the number of hidden nodes between 3 and 11. Nevertheless, we will here impose four procedures (for the 4th procedure see later on): (*i*) first a single hidden node, then (*ii*) hidden nodes from two to twelve will sequentially be tried ‘by hand’ (i.e., program does not allow the imposed number of hidden nodes to be changed), and, finally, (*iii*) the program chooses the number of hidden nodes. To come as close as possible to the MLS results it was decided (*iv*), to compute again the one hidden neuron case where either one or two indices with the lowest sensibility value have

been deleted. In this case, for instance, the number of weights for the 4-1-1 case of T_b is 7, and it equals the number of correlations coefficients from the MLS calculations with six indices.

The results of the five procedures, one MLS and four ANN, given in separate tables, allow a meaningful comparison among them. The MLS procedure optimizes a number of regression parameters equal to the number of variables plus one (the *bias* parameter), which means that a practical comparison between the two methods should only be performed when ANN uses only one hidden neuron. In this case, due to the previous relation, the number of ANN weights equals the number of MLS correlation coefficients plus two. One should expect that with growing number of hidden neurons the model of a property should constantly improve due to the growing number of weights for each variable (akin having a variable with many different regression coefficients). With ANN it is not rare the case that as the model becomes exceedingly good with growing number of weight parameters, and this frequently results in *overfitting* with exceedingly poor externally evaluated values. The choice of training (TR = 80% of the values in Table 3) and test sets (TE = 20% of the values, i.e., the underlined bold values in this same Table) usually avoids overfitting as the network is repeatedly trained for a number of cycles so long as the test error is decreasing, otherwise the training is halted. This method, also known as 'early stopping' procedure [13] avoids the trap that the program will always choose the maximum number of hidden nodes. Actually each property shows an optimal number of nodes which rarely corresponds with its maximum possible.

Results:

Table 4 from ref. 1, collects the MLS results for the six properties. The training set for the *El* property includes pentane and tetrahidrofurane.

Table 5 through 8 collect the various ANN - MLP results for the set of variables (descriptors) of Table 4 or for the alternative (either MMCI or MCI) set of variables obtained with the MLS method. Throughout the Tables 5-8, in the first column are given, as in Table 4, the δ^v type (see Appendix), and the number of networks to train, *Ntr* = 10^3 or 10^5 (when the two numbers rose to similar results *Ntr* = 10^3 was preferred), while the number of networks to retain is always 200. It was not always possible to achieve improvement in modeling with *Ntr* = 10^5 even after three runs (more were not tried), and when there was improvement, it was never sharp excepting two cases as we shall see. The activation functions together with the neuronal architecture are in the second column of Tables 5 – 8. In this column, 3rd line, are given for each property the *number of epochs* for which the ANN-MLP calculation runs even if the actual number of cycles used to train the model might be greater. In fact, as the *number of epochs* is not definitive it cannot be held as an unfailing parameter (it can exceed the given number). In

the third column is the set of variables together with their statistics. Throughout this column in the second line are the sensitivity values, which are the values that are due to the sensitivity analysis that rates the importance of the models' input variables. These r^2 and s, statistics were obtained with the EXCEL spreadsheet plotting the observed *property*, P, vs. the calculated one, P_{clc}, once for the training and test set compounds, $N(a\text{TR} + b\text{TE})$, and the other time for the training {TR} + test {TE} + external evaluation {EV} sets, $N(a\text{TR} + b\text{TE} + c\text{EV})$, where a, b, and c are their number. We remind the reader that the MLS procedure has no test compounds, only training compounds, N(TR). No ANN weights are shown as, for instance, a [5-7-1] network has fifty weights. Furthermore, it is to notice that every time an ANN-MLP runs different weights and sensitivity values are obtained a first non-minor drawback avoided by the MLS procedure.

For comparison purposes it was decided to maintain in the ANN calculations (see Tables 5-8) the same number of outliers excluded throughout the MLS procedure and given in Table 4,where the exclusion was done for residuals greater than 3s. Clearly, such restriction is no more valid throughout ANN tables 5 - 8. In Table 5 are given the ANN results obtained with a single hidden neuron. Following Tables 6 and 7 display the multiple neuron cases: Table 6 with an externally imposed number of hidden neurons that was cycled from 2 to 12, and Table 7 with the number of hidden neurons chosen automatically by the program (between 3 and 11). For *El*, the program sets this number between 3 and 10. For those cases where similar good models are achieved with different sets of hidden nodes, the set with minimal number of nodes was chosen.

Discussion:

For an easier lecture and interpretation the detailed and most important statistical results collected through Tables 4 – 7 are summarized in Table 9. Table 8, illustrates a special case that will be discussed later on. While Tables 4 – 7 collect the detailed information about the modeling of the six properties, and especially about the type of indices, valence deltas, and structure of the ANN computations, Table 9 gives direct information about of the various models. MMCI indices throughout the ANN computations with one hidden neuron, (*ANN* ***1****HN*, Table 5), are important descriptors for flash point, *FP*, and elutropic values, *El*.

As soon as the number of hidden neurons grows either by external choice, *enHN* (Table 6), or by software choice, *snHN* (table 7), MMCIs are optimal descriptors only for Elutropic values (silica), *El*, the property with the lowest number of points. The second thing we notice is that for an optimal modeling the number of hidden neurons (number in bold, 2nd line of the statistical values for each property) that are externally chosen (*ANN enHN*, Table 6) is smaller relatively to the number of hidden neurons chosen by the software (last column, *ANN snHN*, Table 7). In some cases it is much smaller, like for $\boldsymbol{T_b}$ (an

extreme case), ***d***, and ***γ***. Concerning the statistical results we see that *ANN 1HN* (Table 5) improve at the training level (first line) over MLS (Table 4) for T_b, and ***El*** properties, while it stays behind with ***FP***, otherwise results are rather similar. With the whole set of compounds (second line), i.e., with training (and test with ANN) plus evaluated compound *ANN 1HN* (Table 5) calculations improve again for T_b, and ***El***, while they stay behind with ***γ***.

The multiple hidden neuron case shows that, at the training level *ANN enHN* (Table 6), things improve consistently over the two previous cases (*MLS* and *ANN 1HN*) for T_b, ***d***, ***γ***, and ***FP***. For ***El*** there is improvement only relative to the *MLS* (Table 4) case. Results for viscosity, ***η***, are rather similar throughout the three cases. In most cases improvement concerns both the r^2 and the *s* statistics. Concerning the whole set of compounds (tranining plus test and evaluated) statistics improve relatively to the two previous cases (*MLS* and *ANN 1HN*) for T_b, ***γ***, and ***FP***. The advantage of *ANN* over the *MLS* procedure is usually not drastic throughout the six properties. The *ANN snHN* (Table 7, and Table 9, last column) procedure with software chosen number of hidden neurons normally uses more hidden neurons that the previous *ANN enHN* (Table 6, and Table 9, before the last column) one. Actually, it does not achieve any practical improvement. Normally, its statistics are either worse or similar to the *ANN enHN*. This means that if you intend to let the software choose the number of hidden neurons then better you stick to the *MLS* modeling.

In those cases where deletion of two indices resulted in a poor modeling, only an index is deleted. In this last case the number of weights is no more equal (actually is bigger by one) to the number of correlation coefficients of the MLS case. Results are shown in Table 8, and, as the reader can notice, two properties, ***γ***, and ***El***, due to poor modeling do not show up, while for properties ***d***, and ***FP***, only one index has been deleted.

Remind that sensibility values are no absolute values as they normally change from run to run, like the weights, and they are no last word about the importance of an index. The statistics here are usually not as good as in the MLS case (Table 4).

Tables 5 - 8 tell us that there is no fixed preferential value for the parameter *Ntr* (numbers of networks to train), sometimes quite different *Ntr* values give rise to similar statistical values and some other times they give rise to completely different ones. Thus, it is always worth trying several *Ntr* values. Concerning the most used values for δ^v Tables 4 through 7 show that the $\delta^v{}_{ppo}$ configuration is the most used, especially in both *nHN* cases (Tables 6, and 7), which means a strong dependence on the core electrons for heavier atoms (see Appendix). For what concerns the exponent of the fractional term in δ^v (see Appendix) the most used values are 1, -0.5 (strong hydrogen atom dependence), and 50 (no hydrogen atom dependence). The strong

hydrogen dependence of δ^{v} reveals that the hydrogen atoms should not be underestimated.

Table 1. Definition of the Molecular Connectivity Indices (MCI). Replacing δ with δ^{v} and I with S the corresponding valence, χ^{v}, I-State, ψ_I, and E-State, ψ_E, MCIs are obtained.

MCI	*Pseudo-MCI*	*Dual MCI + Δ + Σ*	*Dual pseudo-MCI + $T_{\Sigma/M}$*
$D = \Sigma_i \delta_i$	${}^{S}\psi_I = \Sigma_i I_i$	${}^{0}\chi_d = (-0.5)^N \Pi_i(\delta_i)$	${}^{0}\psi_{Id} = (-0.5)^N \Pi_i(I_i)$
${}^{0}\chi = \Sigma(\delta_i)^{-0.5}$	${}^{0}\psi_I = \Sigma(I_i)^{-0.5}$	${}^{1}\chi_d = (-0.5)^{(N+\mu-1)} \Pi(\delta_i +$	${}^{1}\psi_{Id} = (-0.5)^{(N+\mu-1)} \Pi(I_i +$
${}^{1}\chi = \Sigma(\delta_i\delta_j)^{-0.5}$	${}^{1}\psi_I = \Sigma(I_i I_j)^{-0.5}$	${}^{1}\chi_s = \Pi(\delta_i + \delta_j)^{-0.5}$	${}^{1}\psi_{Is} = \Pi(I_i + I_j)^{-0.5}$
$\chi_t = (\Pi \delta_i)^{-0.5}$	${}^{T}\psi_I = (\Pi I_i)^{-0.5}$	$\Delta = \Sigma_{EA} n_{EA}$, $\Sigma = \Sigma_{EA} <S_{EA}>$	$T_{\Sigma/M} = \Sigma^3 / M^{1.7}$

N is the number of atoms, ij means corresponds to σ bond, μ is the cyclomatic number.

Table 2. Definition of the Mean Molecular Connectivity Indices (MMCI). Replacing δ with δ^{v}, I, and with S the corresponding valence, M^{v}, I-State, M_I, and E-State, M_E, MMCIs are obtained.

${}^{A}M = \Sigma_i \delta_i / n$	${}^{G}M = \Sigma_{ij}(\delta_i\delta_j)^{1/2}$	${}^{H}M = 2\Sigma_{ij} (\delta_i^{-1} + \delta_j^{-1})^{-1}$
${}^{R}M = \Sigma_{ij}[(\delta_i^2 + \delta_j^2)/2]^{1/2}$	${}^{S}M = \Sigma_{ij} (\delta_i^2 + \delta_j^2) / (\delta_i + \delta_j)$	${}^{U}M = \Sigma_{ij}[\delta_i - \delta_j + (\delta_i^2 - 2\delta_i\delta_j + 5\delta_j^2)^{0.5}]/2$
${}^{Ho}M = \Sigma_{ij}(\delta_i^p + \delta_j^p)^{1/p} / 2$	${}^{L}M = \Sigma_{ij}(\delta_i^p + \delta_j^p) / (\delta_i^{p-1} + \delta_j^{p-1})$	${}^{St}M = \Sigma_{ij}[(\delta_i^p - \delta_j^p) / (p\delta_i + p\delta_j)]^{1/(p-1)}$

A: Arithmetic; G: geometric; *H*: harmonic; *R*: root mean square; *S*: symmetric; *U*: unsymmetric; *Ho*: Hölder; *L*: Lehmer; *St*: pseudo-Stolarsky

Table 3. Six properties of organic solvents plus their molar mass M (g·mol^{-1}): ***T_b***, boiling point (K); ***d***, density (at 20°C±5°C relative to water at 4°C, g/cc); ***FP***, flash point (K); ***η***, viscosity (Cpoise, 20°C; [1]at 25°C, [2] at 15°C); ***γ***, surface tension (mN/m at 25°C); and ***El***, Elutropic value (silica).

Solvent	*M*	T_b	*d*	*FP*	*η*	*γ*	*El*
(°)Acetone	58.1	329	0.791	256	0.32	23.46	0.43
(°)Acetonitrile	41.05	355	0.786	278	0.37	28.66	0.50
Benzene	78.1	353	0.84	262	**0.65**	28.22	0.27
Benzonitrile	103.1	461	1.010	344	1.24[1]	38.79	
1-Butanol	74.1	391	0.810	308	2.95	**24.93**	
(°)2-Butanone	72.1	353	0.805	270	0.40	23.97	0.39
Butyl Acetate	116.2	398	0.882	**295**	**0.73**	**24.88**	
CS_2	76.1	319	1.266	**240**	0.37	**31.58**	
CCl_4	153.8	350	1.594		**0.97**	**26.43**	0.14
Cl-Benzene	112.6	405	1.107	296	0.80	32.99	
1Cl-Butane	92.6	351	0.886	267	0.35	23.18	
$CHCl_3$	119.4	334	1.492		0.57	26.67	**0.31**
Cyclohexane	84.2	354	0.779	255	1.00	24.65	0.03
(°)Cyclopentane	70.1	323	0.751	236	0.47	21.88	
1,2-diCl-Benzene	147.0	453	1.306	**338**	1.32		
1,2-diCl-Ethane	98.95	356	1.256	288	0.79	31.86	

diCl-Methane	84.9	313	1.325		0.44	27.20	0.32
N,*N*-diM-Acetamide	87.1	**438**	**0.937**	343			
N,*N*-diM-Formamide	73.1	**426**	**0.944**	**330**	0.92		
1,4-Dioxane	88.1	374	1.034	285	**1.54**	**32.75**	
Ether	74.1	308	0.708	**233**	0.24	16.95	0.29
Ethyl acetate	88.1	350	0.902	270	0.45	23.39	**0.45**
(°)Ethyl alcohol	46.1	351	0.785	281	1.20	21.97	
Heptane	100.2	371	0.684	272		19.65	0.00
Hexane	86.2	**342**	**0.659**	**250**	0.33	17.89	0.00
2-Methoxyethanol	76.1	398	0.965	319	**1.72**	30.84	
(°)Methyl alcohol	32.0	338	0.791	284	0.60	22.07	0.73
4-Me-2-Pentanone	100.2	391	0.800	286			
2-Me-1-Propanol	74.1	381	0.803	310			
2-Me-2-Propanol	74.1	356	0.786	277		19.96	
DMSO	78.1	462	1.101	368	2.24	**42.92**	
(°)Nitromethane	61.0	374	1.127	308	0.67	36.53	
1-Octanol	130.2	469	0.827	354	10.6[2]	27.10	
(°)Pentane	72.15	309	0.626	224	0.23	15.49	0.00*
3-Pentanone	86.1	375	0.853	279		24.74	
(°)1-Propanol	60.1	370	0.804	288	2.26	23.32	
(°)2-Propanol	60.1	356	0.785	295	2.30	20.93	0.63
Pyridine	79.1	388	0.978	293	0.94	36.56	0.55
tetraCl-Ethylene	165.8	394	1.623		0.90		
(°)tetra-Hydrofuran	72.1	340	0.886	256	0.55		0.35*
Toluene	92.1	**384**	**0.867**	277	0.59	27.93	0.22
1,1,2triCl,triFEthane	187.4	**321**	**1.575**		**0.69**		**0.02**
2,2,4-triMe-Pentane	114.2	372	0.692	266	0.50		0.01
o-Xylene	106.2	**417**	**0.870**	**305**	0.81	29.76	
p-Xylene	106.2	411	0.866	300	0.65	28.01	
(°)Acetic acid	60.05	391	1.049			27.10	
Decaline	138.2	**465**	**0.879**				
diBr-Methane	173.8	370	1.542			**39.05**	
1,2-diCl-Ethylen(Z)	96.9	334	1.284				
(°)1,2-diCl-Ethylen(E)	96.9	321	1.255				
1,1-diCl-Ethylen	96.9	**305**	**1.213**				
Dimethoxymethane	76.1	315	0.866				
(°)Dimethylether	46.1	249					
Ethylen Carbonate	88.1	511	1.321				
(°)Formamide	45.0	484	1.133			57.03	
(°)Methylchloride	50.5	249	0.916				
Morpholine	87.1	402	1.005				

Quinoline	129.2	510	1.098	42.59
(°)SO_2	64.1	263	1.434	
2,2-tetraCl-Ethane	167.8	**419**	**1.578**	35.58
tetraMe-Urea	116.2	450	0.969	
triCl-Ethylen	131.4	360	1.476	

(°) externally validated compounds; underlined **bold** values: test compounds used in ANN-MLP calculations, * for this property these two compounds are included in the training set {TR} (Table 4) and training + test sets {TR + TE} (Table 5).

Table 4. Best set of descriptors for the properties of Table 3 with MLS methodology. 1st column: δ^v type for the valence-dependent indices. 2nd column: set of descriptors and their statistical quality.

δ^v - *type*	*Regression equations*
$\delta^v_{po}(1)$	$\boldsymbol{T_b} = 237.5 + 139.1\,{}^0\chi + 24.69 D^v + 527.7\,{}^0\psi_I - 25.91\,{}^1\psi_I - 1500\,{}^0\psi_E + 41.53\,T_{\Sigma/M}$ (24, 31, 3.5, 69, **21**, 222, 10) N(TR) = 45, r^2 = 0.821, s = 22; N(+16EV) = 61, r^2 = 0.792, s = 25
$\delta^v_{ppo}(-0.5)$	$\boldsymbol{d} = 0.733 + 0.024 D^v + 0.211\,{}^0\chi^v + 1.463\,{}^1\chi^v_s - 0.022\,{}^S\psi_E + 0.148\,\Delta$ (0.06, 0.002, 0.02, 0.3, 0.002, 0.01) N(TR) = 45, r^2 = 0.939, s = 0.07; N(+15EV) = 60, r^2 = 0.914, s = 0.08
$\delta^v_{po}(-0.5)$	$\boldsymbol{\gamma} = 8.683 + 0.386 D^v + 397.6\,{}^1\chi^v_s + 151.9\,{}^T\psi_I - 502.4\,{}^1\psi_{Is} + 3.347\,\Delta$ (2.3, 0.05, 57, 36, 90, 0.7) N(TR) = 29, r^2 = 0.835, s = 3.1; N(+10EV) = 39, r^2 = 0.792, s = 3.1
$\delta^v_{ppo}(0.5)$	$\boldsymbol{FP} = 387.1 + 26.99\,{}^H M - 94.38\,{}^H M_I + 33.03\,{}^G M_E + 114.5\,{}^U M_I - 83.10\,{}^{Ho} M_E$ (26, 6.2, 12, 5.2, 13, 11) N(TR) = 29, r^2 = 0.829, s = 16; N(+11EV)= 40, r^2 = 0.764, s = 17
$\delta^v_{po}(-0.5)$	$\boldsymbol{\eta} = -0.216 + 0.001\,{}^1\chi_d + 0.486\,{}^1\psi_I + 2.20\cdot10^{-5}\,{}^1\psi_{Id} - 3.83\cdot10^{-6}\,{}^0\psi_{Ed} + 0.098\,\Sigma$ (0.2, 0.0003, 0.1, $7\cdot10^{-6}$, 10^{-7}, 0.01) N(TR) = 28, r^2 = 0.969, s = 0.4; N(+10EV) = 38, r^2 = 0.939, s = 0.4
$\delta^v_{ppo}(1)$	$\boldsymbol{EI} = 0.018 + 0.181\cdot10^{-3}\,{}^1\chi_d - 0.675\cdot10^{-6}\,{}^1\chi^v_d + 0.003\,{}^0\psi_{Id} + 140.8\,T_{\Sigma/M}$ (0.02, 0.00006, 10^{-7}, 0.0004, 14) N(TR) = 15, r^2 = 0.934, s = 0.06; N(+3EV) = 18, r^2 = 0.931, s = 0.06

*Me = Methyl, THF = tetrahydrofuran, Et = Ethyl.

Table 5. ANN results for the set of descriptors of Table 4 with one hidden neuron. 1st column: the δ^{v}-*type* and the *Ntr* value; 2nd column: ANN-MLP architecture, the abbreviation for the activation functions for the hidden and output layers, the number of epochs, and training and test errors; 3rd column: input indices, their sensitivity value, and statistical parameters for the training plus test, a[*N*(*a*TR + *b*TE)], and plus the evaluation compounds: [*N*(*a*TR + *b*TE + *c*EV)].

δ^{v}-*type*	*ANN-MLP*	(***Variables***) → ***Property***
$\delta^{v}_{po}(1)$ $Ntr = 10^5$	**6 - 1 - 1** (*e*, *l*)* 41 0.005/0.003	$({}^{0}\chi, D^{v}, {}^{0}\psi_{I}, {}^{1}\psi_{I}, {}^{0}\psi_{E}, T_{\Sigma/M}) \rightarrow T_b$ (30.67, 34.22, 41.80, 1.111, 15.76, 2.291) *N*(36TR + 9TE) = 45, r^2 = 0.850, *s* = 21; *N*(+ 16EV) = 61 r^2 = 0.820, *s* = 23 Excluded outlier: dMe-Ether & SO_2 ϵ {EV}
$\delta^{v}_{ppo}(-0.5)$ $Ntr = 10^3$	**5 - 1 - 1** (*t*, *t*) 33 0.002/0.0006	$(D^{v}, {}^{0}\chi^{v}, {}^{1}\chi^{v}_{s}, {}^{S}\psi_{E}, \Delta) \rightarrow d$ (17.99, 8.653, 2.953, 41.31, 12.37) *N*(36TR + 9TE) = 45, r^2 = 0.956, *s* = 0.1; *N*(+ 15EV) = 60, r^2 = 0.930, *s* = 0.1 Excluded outliers: MeCl & MeOH ϵ {EV}
$\delta^{v}_{po}(-0.5)$ $Ntr = 10^5$	**5 - 1 - 1** (*e*, *t*) 27 0.005/0.006	$(D^{v}, {}^{1}\chi^{v}_{s}, {}^{T}\psi_{I}, {}^{1}\psi_{Is}, \Delta) \rightarrow \gamma$ (9.086, 34.48, 34.44, 45.45, 2.328) *N*(22TR + 7TE) = 29, r^2 = 0.841, *s* = 2.8; *N*(+ 10EV) = 39, r^2 = 0.705, *s* = 3.7 Excluded outlier: nitromethane & formamide ϵ {EV}
$\delta^{v}_{ppo}(0.5)$ $Ntr = 10^3$	**5 - 1 - 1** (*e*, *e*) 39 0.009/0.009	$({}^{H}M, {}^{H}M_{I}, {}^{G}M_{E}, {}^{U}M_{I}, {}^{Ho}M_{E}) \rightarrow FP$ (445.1, $1.44\cdot10^6$, $2.65\cdot10^6$, $4.22\cdot10^6$, $17\cdot10^6$) *N*(22TR + 7TE) = 29, r^2 = 0.801, *s* = 16; *N*(+ 11EV) = 40, r^2 = 0.769, *s* = 16 Excluded outliers: 2Me-Butane ϵ {EV}
$\delta^{v}_{po}(-0.5)$ $Ntr = 10^3$	**5 - 1 - 1** (*e*, *l*) 17 0.0006/0.0004	$({}^{1}\chi_{d}, {}^{1}\psi_{I}, {}^{1}\psi_{Id}, {}^{0}\psi_{Ed}, \Sigma) \rightarrow \eta$ (1.982, 1.509, 1.060, 12.04, 3.824) *N*(22TR + 6TE) = 28, r^2 = 0.972, *s* = 0.3; *N*(+ 10EV) = 38, r^2 = 0.942, *s* = 0.4 Excluded outlier: MeOH ϵ {EV}
$\delta^{v}_{ppo}(1)$ $Ntr = 10^3$	**4 - 1 - 1** (*i*, *i*) 20 0.002/0.0003	$({}^{A}M^{v}, {}^{H}M_{E}, {}^{G}M_{E}, {}^{St}M_{I}) \rightarrow EI$ (52.93, 3072, 3020, 27.81) *N*(12TR + 3TE) = 15, r^2 = 0.966, *s* = 0.04; *N*(+ 3EV) = 18, r^2 = 0.955, *s* = 0.04 pentane and THF ϵ {TR}; Excluded. outlier.: MeOH & 2-propanol ϵ {EV}

* Activation functions: *e* = exponential, *i* = identity, *l* = logistic, *t* = tanh. *s* = sin.

Table 6. ANN - MLP results for the set of descriptors of Table 4 with externally imposed number of hidden neurons. 1st column: the δ^v-*type* and the *Ntr* value; 2nd column: ANN-MLP architecture, the abbreviation for the activation functions for the hidden and output layers, the number of epochs, and training and test errors; 3rd column: input indices, their sensitivity value, and statistical parameters for the training plus test, a[*N*(*a*TR + *b*TE)], and plus the evaluation compounds: [*N*(*a*TR + *b*TE + *c*EV)].

δ^v-*type*	*ANN-MLP*	(***Variables***) → ***Property***
	6 - 2 - 1	$({}^0\chi, D^v, {}^0\psi_I, {}^1\psi_I, {}^0\psi_E, T_{\Sigma/M}) \rightarrow T_b$
$\delta^v_{po}(1)$	(*t*, *t*)	(18.17, 50.17, 138.5, 6.414, 93.87, 4.392)
$Ntr = 10^3$	73	*N*(36TR + 9TE) = 45, r^2 = 0.891, *s* = 17; *N*(+ 16EV) = 61 r^2 = 0.871, *s* = 20
	0.004/0.002	Excluded outlier: SO_2 & MeOH ϵ {EV}
	5 - 4 - 1	$(D^v, {}^0\chi^v, {}^1\chi^v_s, {}^S\psi_E, \Delta) \rightarrow d$
$\delta^v_{ppo}(-0.5)$	(*t*, *l*)	(41.54, 29.37, 9.057, 47.73, 29.59)
$Ntr = 10^3$	58	*N*(36TR + 9TE) = 45, r^2 = 0.990, *s* = 0.04; *N*(+ 15EV) = 60, r^2 = 0.966, *s* = 0.1
	0.0004/0.000	Excluded outliers: formamide & MeCl ϵ {EV}.
	5 - 4 - 1	$(D^v, {}^1\chi^v_s, {}^T\psi_I, {}^1\psi_{Is}, \Delta) \rightarrow \gamma$
$\delta^v_{po}(-0.5)$	(*t*, *e*)	(1285, 21.98, 2093, 62687, 5.853)
$Ntr = 10^5$	36	*N*(22TR + 7TE) = 29, r^2 = 0.908, *s* = 2.1; *N*(+ 10EV) = 39, r^2 = 0.871, *s* = 2.4
	0.004/0.002	Excluded outlier: nitromethane & formamide ϵ {EV}
	5 - 5 - 1	$(D, {}^1\psi_{Is}, {}^0\psi_{Ed}, \Delta, T_{\Sigma/M}) \rightarrow FP$
$\delta^v_{po}(1)$	(*t*, *l*)	(8.683, 2.965, 1.212, 5.431, 5.439)
$Ntr = 10^5$	35	*N*(22TR + 7TE) = 29, r^2 = 0.919, *s* = 10; *N*(+ 11EV) = 40, r^2 = 0.860, *s* = 13
	0.003/0.009	Excluded outliers: nitromethane ϵ {EV}
	5 - 3 - 1	$({}^1\chi_d, {}^1\psi_I, {}^1\psi_{Id}, {}^0\psi_{Ed}, \Sigma) \rightarrow \eta$
$\delta^v_{po}(-0.5)$	(*e*, *l*)	(4.609, 5.914, 1.286, 15.86, 6.803)
$Ntr = 10^5$	35	*N*(22TR + 6TE) = 28, r^2 = 0.982, *s* = 0.3; *N*(+ 10EV) = 38, r^2 = 0.975, *s* = 0.3
	0.0003/0.000	Excluded outlier: 2-butanone ϵ {EV}
$\delta^v_{ppo}(1)$	**4 - 2 - 1**	$({}^AM^v, {}^HM_E, {}^GM_E, {}^{St}M_I) \rightarrow EI$
	(*t*, *s*)	(80.08, 3075, 2819, 34.79)
$Ntr = 10^3$	22	*N*(12TR + 3TE) = 15, r^2 = 0.973, *s* = 0.03; *N*(+ 3EV) = 18, r^2 = 0.976, *s* = 0.03
	0.001/0.003	pentane and THF ϵ {TR}; excluded outliers: acetonitrile & 2-propanol ϵ {EV}

Table 7. ANN - MLP results with the number of hidden neurons chosen by Statistica 8. Descriptors are those of Table 4. 1st column: the δ^{v}-*type* and the *Ntr* value; 2nd column: ANN-MLP architecture, the abbreviation for the activation functions for the hidden and output layers, the number of epochs, and training and test errors; 3rd column: input indices, their sensitivity value, and statistical parameters for the training plus test, a[*N*(*a*TR + *b*TE)], and plus the evaluation compounds: [*N*(*a*TR + *b*TE + *c*EV)].

δ^{v}(*type*)	*ANN-MLP*	(***Variables***) → ***Property***
	6 - 11 - 1	$({}^{0}\chi, D^{v}, {}^{0}\psi_{I}, {}^{1}\psi_{I}, {}^{0}\psi_{E}, T_{\Sigma/M}) \rightarrow T_{b}$
$\delta^{v}_{po}(1)$	(*t*, *t*)	(17.98, 45.18, 106.2, 2.556, 72.23, 3.579)
$Ntr = 10^{3}$	39	*N*(36TR + 9TE) = 45, r^{2} = 0.846, *s* = 21; *N*(+ 16EV) = 61 r^{2} = 0.826, *s* = 24
	0.005/0.005	Excluded outlier: MeOH & SO_2 ∈ {EV}
	5 - 8 - 1	$(D^{v}, {}^{0}\chi^{v}, {}^{1}\chi^{v}_{s}, {}^{S}\psi_{E}, \Delta) \rightarrow d$
$\delta^{v}_{ppo}(-0.5)$	(*t*, *I*)	(20.47, 8.414, 4.606, 49.56, 19.77)
$Ntr = 10^{5}$	18	*N*(36TR + 9TE) = 45, r^{2} = 0.970, *s* = 0.05; *N*(+ 15EV) = 60, r^{2} = 0.938, *s* = 0.07
	0.001/0.001	Excluded outliers: MeCl & MeOH ∈ {EV}
	5 - 10 - 1	$(D^{v}, {}^{1}\chi^{v}_{s}, {}^{T}\psi_{I}, {}^{1}\psi_{Is}, \Delta) \rightarrow \gamma$
$\delta^{v}_{po}(-0.5)$	(*I*, *s*)	(18.16, 81.96, 74.19, 173.8, 2.809)
$Ntr = 10^{3}$	42	*N*(22TR + 7TE) = 29, r^{2} = 0.890, *s* = 2.3; *N*(+ 10EV) = 39, r^{2} = 0.851, *s* = 2.6
	0.004/0.002	Excluded outlier: nitromethane & formamide ∈ {EV}
	5 - 4 - 1	$(D, {}^{1}\psi_{Is}, {}^{0}\psi_{Ed}, \Delta, T_{\Sigma/M}) \rightarrow FP$
$\delta^{v}_{po}(1)$	(*I*, *I*)	(6.663, 2.542, 1.105, 4.616, 3.220)
$Ntr = 10^{5}$	81	*N*(22TR + 7TE) = 29, r^{2} = 0.899, *s* = 11; *N*(+ 11EV) = 40, r^{2} = 0.840, *s* = 14
	0.003/0.01	Excluded outliers: 2Me-Butane ∈ {EV}
	5 - 3 - 1	$({}^{1}\chi_{d}, {}^{1}\psi_{I}, {}^{1}\psi_{Id}, {}^{0}\psi_{Ed}, \Sigma) \rightarrow \eta$
$\delta^{v}_{po}(-0.5)$	(*e*, *I*)	(6.071, 4.640, 1.164, 14.16, 7.089)
$Ntr = 10^{3}$	26	*N*(22TR + 6TE) = 28, r^{2} = 0.981, *s* = 0.3; *N*(+ 10EV) = 38, r^{2} = 0.974, *s* = 0.3
	0.0003/0.0003	Excluded outlier: 2-butanone ∈ {EV}
	4 - 5 - 1	$({}^{A}M^{v}, {}^{H}M_{E}, {}^{G}M_{E}, {}^{St}M_{I}) \rightarrow EI$
$\delta^{v}_{ppo}(1)$	(*t*, *t*)	(66.31, 355.9, 331.7, 27.55)
$Ntr = 10^{5}$	49	*N*(12TR + 3TE) = 15, r^{2} = 0.973, *s* = 0.03; *N*(+ 3EV) = 18, r^{2} = 0.973, *s* = 0.03
	0.002/0.001	pentane and THF ∈ {TR} and excluded MeOH & 2-propanol ∈ {EV}

Table 8. ANN results for the set of descriptors of Table 4 with only one hidden neuron but where either one or two indices have been left out, usually, those with lowest sensitivity values in Table 5. For the structure of this table see Table 5. Only satisfactory results are shown here.

δ^v-type	ANN-MLP	(***Variables***) → ***Property***
	4 - 1 - 1	$({}^0\chi, D^v, {}^0\psi_I, {}^0\psi_E) \rightarrow T_b$
$\delta^v_{po}(1)$	(*e*, e)	(816.3, 863.6, 110900, 7016972)
$Ntr = 10^5$	25	N(36TR + 9TE) = 45, r^2 = 0.758, s = 26; N(+ 16EV) = 61 r^2 = 0.714, s = 29
	0.008/0.008	Excluded outlier: dMe-Ether & SO_2 ϵ {EV}
	4 - 1 - 1	$(D^v, {}^0\chi^v, {}^s\psi_E, \Delta) \rightarrow d$
$\delta^v_{ppo}(-0.5)$	(*I*, *t*)	(11.01, 7.934, 28.40, 4.905)
$Ntr = 10^3$	17	N(36TR + 9TE) = 45, r^2 = 0.917, s = 0.1; N(+ 15EV) = 60, r^2 = 0.895, s = 0.1
	0.004/0.002	Excluded outliers: SO_2 & Formamide ϵ {EV}
	4 - 1 - 1	$({}^HM_I, {}^GM_E, {}^UM_I, {}^{Ho}M_E) \rightarrow FP$
$\delta^v_{ppo}(0.5)$	(*i*, *I*)	(10.65, 14.68, 15.90, 12.16)
$Ntr = 10^5$	26	N(22TR + 7TE) = 29, r^2 = 0.719, s = 19; N(+ 11EV) = 40, r^2 = 0.702, s = 18
	0.01/0.02	Excluded outliers: 2Me-Butane ϵ {EV}
	3 - 1 - 1	$({}^1\chi_d, {}^0\psi_{Ed}, \Sigma) \rightarrow \eta$
$\delta^v_{po}(- 0.5)$	(*t*, *i*)	(1.603, 15.54, 10.70)
$Ntr = 10^3$	67	N(22TR + 6TE) = 28, r^2 = 0.965, s = 0.4; N(+ 10EV) = 38, r^2 = 0.917, s = 0.5
	0.0007/0.0003	Excluded outlier: MeOH ϵ {EV}

Table 9. Statistical, N/r^2(2nd decimal figure)/*s*, results for the six properties from Tables 4 to 7. **2nd column:** *MLS*, results, **3rd column**: *ANN* with one hidden neuron (*ANN **1**HN*) results, **4th column**: *ANN* with externally chosen number of hidden neurons (*ANN enHN*) results, **5th column**: *ANN* with software chosen number of hidden neurons (*ANN snHN*) results. First line shows the statistical results for the training (MLS) and train plus test (ANN) compounds, the second line shows the overall statistical results inclusive the evaluated compounds. ***M*** (bold and italics) stands for MMCIs (otherwise they are MCIs). In the last two columns are also given the number of hidden neurons (second line, italics and bold).

P	*MLS* (Table 4)	*ANN **1**HN* (Table 5)	*ANN enHN* (Table 6)	*ANN snHN* (Table 7)
T_b	45 / 0.82 / 22	45 / 0.85 / 21	45 / 0.89 / 17	45 / 0.85 / 21
	61/ 0.79 / 25	61 / 82 / 23	***2*** / 61 / 87 / 20	***11*** / 61 / 0.83 / 24
d	45/ 0.94 / 0.07	45 / 0.96 / 0.1	45 / 0.99 / 0.04	45 / 0.97 / 0.05
	60 / 0.91 / 0.08	60/ 0.93 / 0.1	*4* / 60 / 0.97 / 0.1	*8* / 60 / / 0.94 / 0.07
γ	29 / 0.84 / 3.1	29 / 0.84 / 2.8	29 / 0.91 / 2.1	29 / 0.89 / 2.3
	39 / 0.79 / 3.1	39 / 0.71 / 3.7	***4*** / 39 / 0.87 / 2.4	***10*** / 39 / 0.85 / 2.6
FP	***M*** / 29 / 0.83 / 16	***M*** / 29 / 0.80 / 16	29 / 0.92 / 10	29 / 0.90 / 11
	40 / 0.76 / 17	40 / 0.77 / 16	***5*** / 40 / 0.86 / 13	***4*** / 40 / 0.84 / 14
η	28 / 0.97 / 0.4	28 / 0.97 / 0.3	28 / 0.98 / 0.3	28 / 0.98 / 0.3
	38 / 0.94 / 0.4	38 / 0.94 / 0.4	***3*** / 38 / 0.98 / 0.3	***3***/ 38 / 0.97 / 0.3
El	15 / 0.93 / 0.06	***M*** / 15 / 0.97 / 0.04	***M*** / 15 / 0.97 / 0.03	***M*** / 15 / 0.97 / 0.03
	18 / 0.93 / 0.06	18 / 0.96 / 0.04	***2*** / 18 / 0.98 / 0.03	***5*** / 18 / 0.97 / 0.03

Conclusions:

The first interesting result of the present ANN computations is that they prefer MCIs instead of MMCIs, especially with properties with relatively large number of points. In fact, only *El*, with minimal number of points is advantageously described by MMCI when ANN with more than one hidden neuron is used. The second result being that it is better to impose from outside the number of hidden neurons. The third result being that it is better to run ANNs using quite different numbers of networks to train, *Ntr*. The fourth and the more interesting result being that normally ANN improve over the MLS calculations, but also that in many cases this improvement is not striking.

It should be remembered that MLS is anyway used to derive the best set of descriptors that are passed over to the ANNs, and that its statistical results are definitive, i.e., no matter how many times you repeat the calculations with the same indices you will obtain always the same results at every statistical level. ANN results are instead unsystematic and non-reproducible as the weights of the ANN

computations start from random values and the minimization procedure usually ends up with different values from run to run. This fuzzy character has nevertheless a positive side as if ANN computations are run over and over again the probability to end up with a quite good result is increasing. *ANN* results obtained with one hidden neuron either with the full set of descriptors (Table 5), or with a reduced set of descriptors, like in Table 8, if on one side they confirm the validity of the MLS calculations on the other side they leave open the possibility that somewhere there are ANN calculations that improve over them.

Anyway, (*i*) before throwing away a bad model for the training plus test compounds with ANN computations think it twice because they could hide a very good model for the evaluated compounds, and (*ii*) do not throw away the hydrogen atoms in calculations with MCIs or MMCIs as in many cases they are of good help.

References:

1. L. Pogliani and J.V.de Julián-Ortiz, *Int. J. Chem. Mod*. 2014, 6, 241-254.

2. M. Randić, *J.Am.Chem.Soc*. 1975, 97, 6609-6615.

3. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*, Wiley, New York, 1986.

4. Kier, L.B., Hall, L.H., *Molecular Structure Description. The Electrotopological State*, Academic Press, New York, 1999.

5. R. Todeschini and V. Consonni, *Molecular Descriptors for Chemoinformatics*, 2nd edn, Wiley-VCH, Weinheim, 2000.

6. L. Pogliani, *Chem.Rev*. 2000, *100*, 3827-3858.

7. R. García-Domenech, J. Gálvez, J. V. de Julián-Ortiz and L. Pogliani, *Chem. Rev.* 2008, 108, 1127-1169.

8. L. Pogliani and J. V. de Julián-Ortiz, *RSC Advances* 2013, 3, 14710-14721.

9. L. Pogliani, J. V. de Julián-Ortiz, *RSC Advances* 2014, 4, 44733-44740.

10. J. G. Topliss and R. J. Costello, *J. Med. Chem*. 1972, 15, 1066-1069.

11. J. Zupan and J. Gasteiger, *Neural Networks in Chemistry and Drug Design*: *An Introduction*, 2nd Edition, Wiley-VCH, Weinheim, 1999.

12. E. Castillo, B. Guijarro-Berdiñas, O. Fontenla-Romero and A Alonso-Betanzos, *J. of Machine Learning Research* 2006, 7, 1159–1182.

13. D. J. Livingstone, D.T. Manallack and I.V. Tetko, J. *Comput.-Aided Mol. Design*.1997, 11, 135-142.