Use of Artificial Neural Network-Group Contribution Method to Determine Surface Tension of Pure Compounds

Farhad Gharagheizi,[†] Ali Eslamimanesh,[‡] Amir H. Mohammadi,[*,‡,§] and Dominique Richon[‡]

[†]Saman Energy Giti, 3331619636 Tehran, Iran

[‡]MINES ParisTech, CEP/TEP - Centre Énergétique et Procédés, 35 Rue Saint Honoré, 77305 Fontainebleau, France

[§]Thermodynamics Research Unit, School of Chemical Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa

Ⓢ Supporting Information

ABSTRACT: This work aims at applying an artificial neural network-group contribution method to represent/predict the surface tension of pure chemical compounds at different temperatures and atmospheric pressure. To propose a comprehensive, reliable, and predictive tool, about 4700 data belonging to experimental surface tension values of 752 chemical compounds at different temperatures and atmospheric pressure have been studied. The investigated compounds belong to 78 chemical families containing 151 functional groups (group contributions), which include organic and inorganic liquids. Using this dedicated strategy, we obtain satisfactory results quantified by the following statistical parameters: absolute average deviations of the represented/predicted properties from existing experimental values, 1.7 %, and squared correlation coefficient, 0.997.

1. INTRODUCTION

Effects of different physical forces on fluid phase equilibria have generated number of discussions in the past century.[1] Surface tension is among these forces, which exists at the interface of fluid phases.[1−3] There are unequal asymmetric forces acting upon a molecule, which are zero at equilibrium.[4] At low gas densities, the molecules experience a sidewise and toward the bulk liquid, meanwhile they are attracted a little in the direction of the bulk gas. These attractive forces tend to pull the surface toward the bulk liquid phase. Therefore, the surface layer is in tension, and at equilibrium, it minimizes its area, which depends on the mass of material, container restraints, and external forces.[3,4] As a matter of fact, accurate determination of the tension at the surface of a liquid is significant for investigation of intermolecular forces.[4]

This molecular tension at the surface is quantitatively expressed as surface (interfacial) tension, which refers to the force exerted at the interface per unit length.[2] Surface forces affect the onset of formation of new phases and are significant in multiphase flow especially in hydrocarbon reservoirs during production and in pipelines during transportation.[2]

One of the applications of this property is to determine the capillary pressure, which is used to investigate the effects of surface forces on fluid distribution within a reservoir.[2] Furthermore, the relative permeability of the fluids that is a significant factor for describing the fluids flow and phase behavior (in a dynamic way) in the hydrocarbon reservoirs is related to the interfacial tension.[2] Of particular interest are the effects of the values of interfacial tensions of the gas condensates on the condensate recovery in the case of retrograde condensation in gas condensate reservoirs.[2]

Regarding the preceding significance of surface tension property, several calculation/estimation methods have been so far proposed for this purpose. In 1923, Macleod[5] presented an empirical equation to correlate the experimental values of surface tension based on the density difference between the liquid and vapor of a chemical compound in equilibrium with each other at a given temperature and a constant characteristic of the liquid phase. Later, Sugden[6,7] reported the constant characteristic of the Macleod's correlation[5] to be a function of molecular weight and another parameter of a compound called the "parachor" as follows:

$$\sigma^{(1/4)} = (P)(\rho_L - \rho_V)/M \tag{1}$$

where σ is the surface tension in (dyn/cm), P denotes the parachor parameter, ρ is the density in (g/cm^3), M is the molecular weight, and subscripts L and V refer to the liquid and vapor phases, respectively. Balasubrahmanyam[1] stated that parachor is a number that represents the molar volume of a compound when the temperature is such that its surface tension is unity.[1] In other words, the parachor parameter has been related to the critical volume of a substance has been related to the critical volume. Therefore, there are unique values of this parameter for each chemical compound.[2] Bayliss[8] has calculated the parachor values by fitting the experimental data of n-paraffins parachors using the least-squares method. A correlation based on molecular weight has been presented by Schechter and Guo[9] and Baker and Swendloff[10] for the evaluation of the parachors of n-paraffins.

However, the parachor parameter can be related to the critical temperature and molar volume of compounds as follows:[11]

$$P = 0.324 T_c^{1/4} v_c^{7/8} \tag{2}$$

where T is temperature in K, v is molar volume in m^3/kmol, and

subscript c denotes the critical value. Another approach has been presented by Quayle,[12] who reported group contributions for calculation of parachor parameters. However, the reported group contributions are incomplete and many functional groups are not represented; that is, the model is not applied to many chemical compounds.[3]

Although the method suggested by Macleod,[5] which relates the surface tension to the densities has some drawbacks, several authors have used this approach to evaluate the surface tensions of various compounds[13−19] mainly due to its simplicity. The main disadvantages of applying this relation are as follows:

1. The values of the parachors are not always available experimentally and estimation techniques have not been developed for many chemical families,
2. The absolute average deviations (AAD) of the calculated/estimated surface tensions of different substances from experimental data are too large for complex chemical structures.[3,4] The obtained results for different kinds of chemical compounds[3] show 8.6 % absolute average deviation from experimental values.

Corresponding state principles[20] have been also developed for correlating the surface tension data of pure compounds.[4,19−27] For instance, Brock and Bird[19] applied the reduced temperature, the Riedel[28] parameter at the critical point, critical temperature, and critical pressure to correlate the surface tension experimental data for nonpolar liquids. This correlation also results in high AAD (14 %) from the same experimental surface tension values, upon which the Macleod's correlation[4] was tested. Another attempt has been made by Curl and Pitzer[21] and Pitzer[22] that proposed critical temperature, pressure, and acentric factor (ω) as the parameters of a correlation for calculation of the surface tensions. The obtained results using this correlation show AAD of 17 %.[3]

Two-reference corresponding states methods have been first applied by Rice and Teja,[24] who used critical temperature and volume to derive a correlation, and later Zuo and Stenby,[23] who reported critical pressure and temperature to calculate surface tensions. These equations do not lead to satisfactory results for chemical compounds with strong hydrogen-bonding forces.[3] To overcome these shortcomings, Sastri and Rao[25] evaluated the surface tensions by a correlation based on critical pressure and temperature, normal boiling point, reduced temperature, and reduced boiling temperature. Their proposed correlation leads to AAD (4 %) in comparison with the surface tensions experimental values of 30 pure compounds.[3]

Freita et al.[26] proposed a different approach, in which a linear solvation free energy relationship (LSER) on the basis of solute parameters for the organic compounds has been applied for estimation of the interfacial tensions of organic liquids. Apart form that, Kavun and co-workers[27] proposed a simple quantitative regression equation using quantitative structure property relationship (QSPR), the graph theory, and several techniques of physical and theoretical organic chemistry to acceptably predict the values of surface tension of liquid organic compounds from various chemical families.

The most accurate equation for representation/prediction of the surface tensions of pure liquids has been proposed by Escobedo and Mansoori,[4] who have derived an expression based on statistical-mechanic and corresponding state principle. Their results show AAD slightly higher than 1 % and around 2.6 % for representation and prediction of surface tensions of 94 various organic compounds, respectively.[4] It should be pointed out that for presenting a precise comparison among the described methods, the same data sets should be applied to check their accuracy, reliability and comprehensiveness.

However, there is still a need for presentation of more comprehensive and accurate methods to determine surface tensions of large group of chemical compounds containing organic and inorganic liquids. In this communication, we propose a new approach based on the artificial neural network-group contribution (ANN-GC) method to represent/predict the surface tension of pure chemical compounds at different temperatures and atmospheric pressure.

2. EXPERIMENTAL DATA AND MATHEMATICAL METHODS

2.1. Experimental Data. The accuracy and reliability of models for representation/prediction of physical properties, especially those dealing with large number of experimental data, directly depends on the quality and comprehensiveness of the applied data set for its development.[29] These characteristics of such a model include both diversity in the investigated chemical families and the number of pure compounds available in the data set. In this work, we used DIPPR 801 database,[30] which is one of the best sources of physical property data for pure compounds that generated based on 23000 scientific sources. The surface tension values of 752 chemical species from various chemical families at different temperatures (4672 data points) have been considered for the calculation procedure. All of the data points have been evaluated by the DIPPR 801 project[30] for organic and inorganic compounds. The references of the data can be found in the Supporting Information.

2.2. Development of New Group Contributions. Having defined the database, the chemical structures of all of the studied compounds are analyzed with much attention using an algorithm comparing the chemical groups to define the most efficient contributions for evaluation of the surface tension. Consequently, 151 functional groups have been found to be more efficient for representation/prediction of the surface tensions of pure compounds. The functional groups used in this study are presented in Table 1. Moreover, the table of their numbers of occurrences in the investigated compounds is presented as Supporting Information. These chemical groups along with the temperature are used as the proposed model parameters.

2.3. Optimization of Group Contributions using Artificial Neural Network. The first calculation step and perhaps the most significant one, is to search for a relationship between the chemical functional groups and the desired physical properties. The simplest method for this purpose is assumption of existence of a multilinear relationship between these groups and the desired property (here the surface tensions).[31] This technique is a similar method used in the most of classical group contribution methods.[29] Several calculations show that application of the mentioned methodology for the current problem brings about poor results. Consequently, nonlinear mathematical method of artificial neural network (ANN) is preferred and investigated. ANN have been applied to various scientific and engineering applications,[29−66] e.g., calculations/estimations of physical and chemical properties of different pure compounds,[46−59,64,65] and phase behavior predictions of complex semiclathrate systems.[45] The theoretical explanations about neural networks have been well presented before.[65] Using the ANN toolbox of the

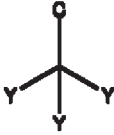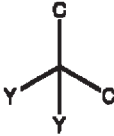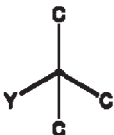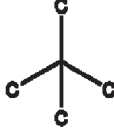## Table 1. Functional Groups Used to Develop the Model[*]
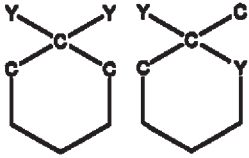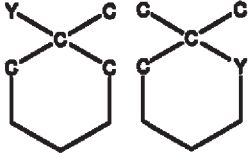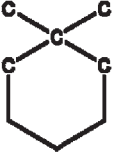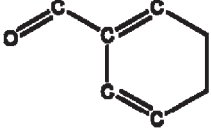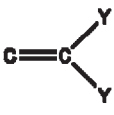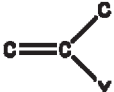
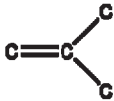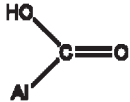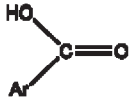| No.	Functional Groups	Comments
1	ST001	terminal primary C(sp3) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH2, etc.)
2	ST002	total secondary C(sp3) Y = H or any heteroatom
3	ST003	total tertiary C(sp3) Y = H or any heteroatom
4	ST004	total quaternary C(sp3)
5	ST005	ring secondary C(sp3) Y = H or any heteroatom
6	ST006	ring tertiary C(sp3) Y = H or any heteroatom
7	ST007	ring quaternary C(sp3)
8	ST008	non-aromatic conjugated C(sp2)
9	ST009	terminal primary C(sp2) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH2, etc.)
10	ST010	aliphatic secondary C(sp2) Y = H or any heteroatom

Table 1. Continued
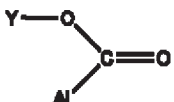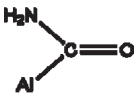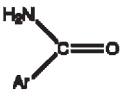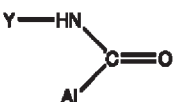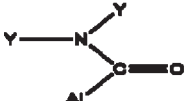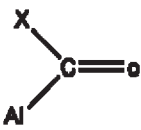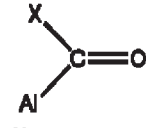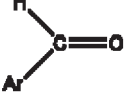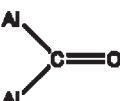
No.		Functional Groups	Comments
11	ST011		aliphatic tertiary C(sp2)
12	ST012		terminal C(sp) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH2, etc.)
13	ST013		carboxylic acids (aliphatic)
14	ST014		carboxylic acids (aromatic)
15	ST015		esters (aliphatic) Y = Ar or Al (not H) Al = H or aliphatic group linked through C
16	ST016		esters (aromatic) Y = Al or Ar
17	ST017		primary amides (aliphatic) Al = H or aliphatic group linked through C
18	ST018		secondary amides (aliphatic) Y = Ar or Al (not H, not C = O) Al = H or aliphatic group linked through C
19	ST019		secondary amides (aliphatic) Y = Ar or Al (not H, not C = O) Al = H or aliphatic group linked through C
20	ST020		tertiary amides (aliphatic) Y = Ar or Al (not H, not C = O) Al = H or aliphatic group linked through C
21	ST021		acyl halogenides (aliphatic)
22	ST022		acyl halogenides (aliphatic)
23	ST023		aldehydes (aromatic)
24	ST024		ketones (aliphatic)

Table 1. Continued
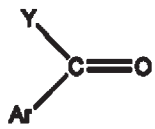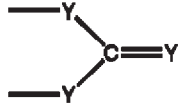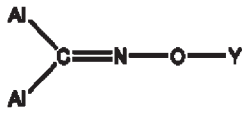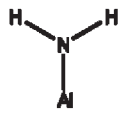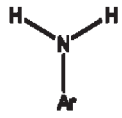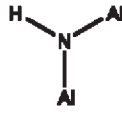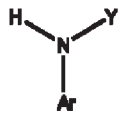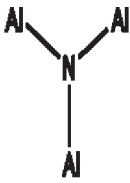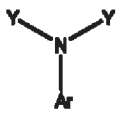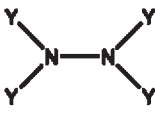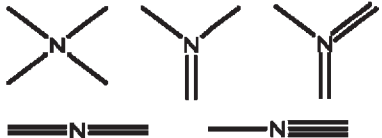
No.		Functional Groups	Comments
25	ST025		ketones (aromatic) Y = Al or Ar
26	ST026		carbonate (-thio) derivatives (Y = O or S)
27	ST027		oximes (aliphatic) Y = H, Ar or Al
28	ST028		primary amines (aliphatic) Al = aliphatic group linked through C (not C = O)
29	ST029		primary amines (aromatic)
30	ST030		secondary amines (aliphatic) Al = aliphatic group linked through C (not C = O)
31	ST031		secondary amines (aromatic) Y = Ar or Al (not C = O)
32	ST032		tertiary amines (aliphatic) Al = aliphatic group linked through C (not C = O
33	ST033		tertiary amines (aromatic) Y = Ar or Al (not C = O)
34	ST034		N hydrazines Y = C or H
35	ST035		nitriles (aliphatic)
36	ST036		nitriles (aromatic)
37	ST037		positive charged N

Table 1. Continued

No.		Functional Groups	Comments
38	ST038	Al—N≡O	nitroso groups (aliphatic)
39	ST039	Al—N(=O)(O)	nitro groups (aliphatic) Al = H or aliphatic group linked through carbon
40	ST040	Ar—N(=O)(O)	nitro groups (aromatic) Al = aromatic group linked through carbon
41	ST041	Al—O—H	hydroxyl groups Al = aliphatic group linked through any atom
42	ST042	Ar—O—H	aromatic hydroxyls Ar = aromatic group linked through any atom
43	ST043	o—C(H)(H)—O—H	primary alcohols
44	ST044	C—C(C)(H)—O—H	secondary alcohols
45	ST045	C—C(C)(C)—O—H	tertiary alcohols
46	ST046	Al—O—Al	ethers (aliphatic) Al = aliphatic group linked through C (not C = O, not C # N)
47	ST047	Ar—O—Y	ethers (aromatic) Y = Ar or Al (not C = O, not C # N)
48	ST048	(Y=)C—O—C(=Y)	anhydrides (thio-) Y = O or S
49	ST049	H-O-H	water molecules
50	ST050	C—S—H	thiols
51	ST051	C—S—C	sulfides
52	ST052	C,C—S=O ; C=S=O ; S=S=O	sulfoxides
53	ST053	O=S=O ; O=S(C)=O ; O=S(S)=O	sulfones

Table 1. Continued

No.		Functional Groups	Comments
54	ST054		sulfonic (thio-/dithio-) acids Y = O or S
55	ST055		sulfites (thio-/ dithio-) Y = O or S
56	ST056		sulfates (thio- / dithio-) (Y = O or S)
57	ST057		phosphanes Y = H, C, halogens
58	ST058		phosphoranes / thiophosphoranes Y = O or S
59	ST059		CH2RX
60	ST060		CHR2X
61	ST061		CR3X
62	ST062		R=CHX
63	ST063		R=CRX
64	ST064		CHRX2
65	ST065		CR2X2

Table 1. Continued
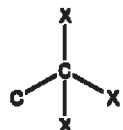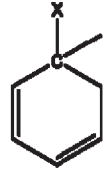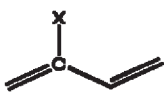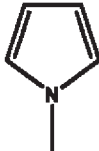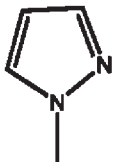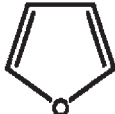
No.		Functional Groups	Comments
66	ST066		CRX3
67	ST067		X on aromatic ring
68	ST068		X on ring C(sp3)
69	ST069		X on exo-conjugated C
70	ST070		Aziridines
71	ST071		Oxiranes
72	ST072		Thioethanes
73	ST073		Pyrrolidines
74	ST074		Thiranes
75	ST075		tetrahydro-Thiophenes
76	ST076		Pyrroles
77	ST077		Pyrazoles
78	ST078		Furanes

Table 1. Continued
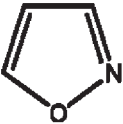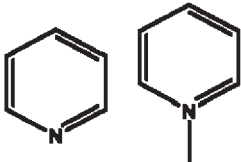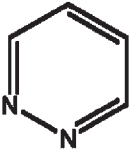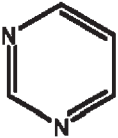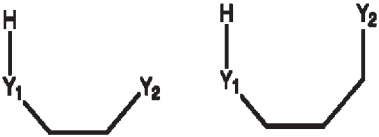
No.		Functional Groups	Comments
79	ST079		Thiophenes
80	ST080		Isoxazoles
81	ST081		Pyridines
82	ST082		Pyridazines
83	ST083		Pyrimidines
84	ST084	Sum of the hydrogens linked to all of the Os and Ns in the molecule	donor atoms for H-bonds (N and O)
85	ST085	Total Ns, Os and Fs in the molecule, excluding N with a formal positive charge, higher oxidation states and pyrrolyl form of N	acceptor atoms for H-bonds (N, O, F)
86	ST086		
87	ST087	CH3R / CH4	
88	ST088	CH3X	
89	ST089	CH2RX	
90	ST090	CH2X2	
91	ST091	CHR2X	
92	ST092	CHRX2	
93	ST093	CHX3	
94	ST094	CR3X	
95	ST095	CX4	
96	ST096	=CH2	
97	ST097	=CHX	
98	ST098	=CRX	
99	ST099	R--CR--R	
100	ST100	R--CX--R	

Table 1. Continued

No.		Functional Groups	Comments
101	ST101	R--CH--X	
102	ST102	R--CR--X	
103	ST103	R--CH..X	
104	ST104	R--CR..X	
105	ST105	Al-CH=X	
106	ST106	R-C(=X)-X / R-C#X / X=C=X	
107	ST107	X-C(=X)-X	
108	ST108	H^a attached to C^1(sp3) / C^0(sp2)	
109	ST109	H^a attached to C^2(sp3) / C^1(sp2) / C^0(sp)	
110	ST110	H^a attached to C^3(sp3) / C^2(sp2) / C^3(sp2) / C^3(sp)	
111	ST111	H attached to alpha-C^b	
112	ST112	H^a attached to C^0(sp3) with 1X attached to next C	
113	ST113	H^a attached to C^0(sp3) with 2X attached to next C	
114	ST114	H^a attached to C^0(sp3) with 3X attached to next C	
115	ST115	alcohol	
116	ST116	phenol / enol / carboxyl OH	
117	ST117	=O	
118	ST118	Al-O-Al	
119	ST119	Al-O-Ar / Ar-O-Ar / R..O..R / R-O-C=X	
120	ST120	O- (negatively charged)	
121	ST121	Any-Se-Any	
122	ST122	Al3-N	
123	ST123	Ar-NH2 / X-NH2	
124	ST124	Ar-NH-Al	
125	ST125	RCO-N< / >N-X=X	
126	ST126	Ar2NH / Ar3N / Ar2N-Al / R..N..R^c	
127	ST127	Al-NO2	
128	ST128	Ar-N=X / X-N=X	
129	ST129	N+ (positively charged)	
130	ST130	F^a attached to C^1(sp3)	
131	ST131	F^a attached to C^1(sp2)	
132	ST132	F^a attached to C^2(sp2)-C^4(sp2) / C^1(sp) / C^4(sp3) / X	
133	ST133	Cl^a attached to C^1(sp3)	
134	ST134	Cl^a attached to C^2(sp3)	
135	ST135	Cl^a attached to C^3(sp3)	
136	ST136	Cl^a attached to C^1(sp2)	
137	ST137	Cl^a attached to C^2(sp2)-C^4(sp2) / C^1(sp) / C^4(sp3) / X	
138	ST138	Br^a attached to C^1(sp3)	
139	ST139	Br^a attached to C^2(sp3)	
140	ST140	Br^a attached to C^3(sp3)	
141	ST141	Br^a attached to C^1(sp2)	
142	ST142	Br^a attached to C^2(sp2)-C^4(sp2) / C^1(sp) / C^4(sp3) / X	
143	ST143	I^a attached to C^1(sp3)	
144	ST144	I^a attached to C^1(sp2)	
145	ST145	chloride ion	
146	ST146	R2S / RS-SR	
147	ST147	R=S	
148	ST148	R-SO-R	
149	ST149	R-SO2-R	
150	ST150	PX3 (phosphite)	
151	ST151	PR3 (phosphine)	

*R represents any group linked through carbon. X represents any electronegative atom (O, N, S, P, Se, halogens). Al and Ar represent aliphatic and aromatic groups, respectively. ⁼ Represents a double bond. # Represents a triple bond. ¯ Represents an aromatic bond as in benzene or delocalized bonds such as the N—O bond in a nitro group. . Represents aromatic single bonds as the C—N bond in pyrrole. ^{a} The superscript represents the formal oxidation number.
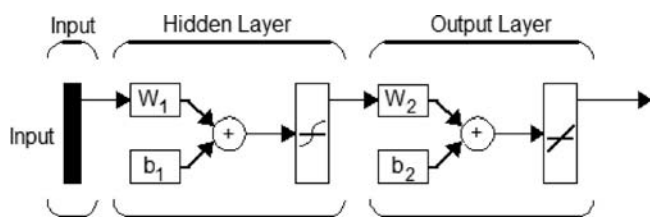
Figure 1. Schematic structure of the three-layer FFANN used in this study. W, weight; b, bias.

MATLAB software (Mathworks Inc.), a three layer feed forward artificial neural network (FFANN) has been developed for the problem. The typical structure of a three layer FFANN is schematically presented in Figure 1. The capabilities of this kind of ANNs have been shown elsewhere.[29−66]

All of the functional groups and also the properties values of pure compounds are normalized between −1 and +1 to decrease computational errors. This can be performed using maximum and minimum values of each functional group for input data and using maximum and minimum values of surface tensions for output parameters. Due to the fact that we are faced with a large range of surface tension values for different compounds, these values are generally normalized between −1 and +1 to prevent truncation errors. In addition, this procedure, which is done in optimization process, is performed to obtain the parameters of the neural networks (W_1, W_2, b_1, and b_2 as shown in Figure 1), and it has no effects on the model results. Later, these values are again changed to the original surface tension values, which are finally used as the inputs and reported as outputs of the developed model. In the next step, the database is divided into three subdata sets including the "Training" set, the "Validation" set, and the "Test" set. In this work, the "Training" set is used to generate the ANN structure, the "Validation (optimization)" set is applied for optimization of the model, and the "Test (prediction)" set is used to investigate the prediction capability and validity of the obtained model. The process of division of database into three subdata sets is performed randomly. For this purpose, about 80 %, 10 %, and 10 % of the main data set are randomly selected for the "Training" set (about 3738 surface tension data), the "Validation" set (467 surface tension data), and the "Test" set (467 surface tension data). The effect of the percent allocation of the three subdata sets from the database on the accuracy of the ANN model has been studied elsewhere.[66]

Developing an ANN method is determination of the weight matrices and bias vectors.[26] As shown in the Figure 1, there are two weight matrices and two bias vectors in a three layer FFANN: W_1 and W_2 and b_1 and b_2.[29−66] These parameters should be obtained by minimization of an objective function. The objective function used in this study is sum of squares of errors between the outputs of the ANN (represented/predicted properties) and the target values (experimental surface tensions). This minimization is performed by Levenberg−Marquardt (LM)[62] optimization strategy. There are also more accurate optimization methods other than this algorithm; however, they need much more convergence time. In other words, the more accurate optimization, the more time is needed for the algorithm to converge to the global optimum. The LM[65] is most-widely used.[29−66]

In most cases, the number of neurons in the hidden layer (n) is fixed. Therefore, the main goal is to produce an ANN model,
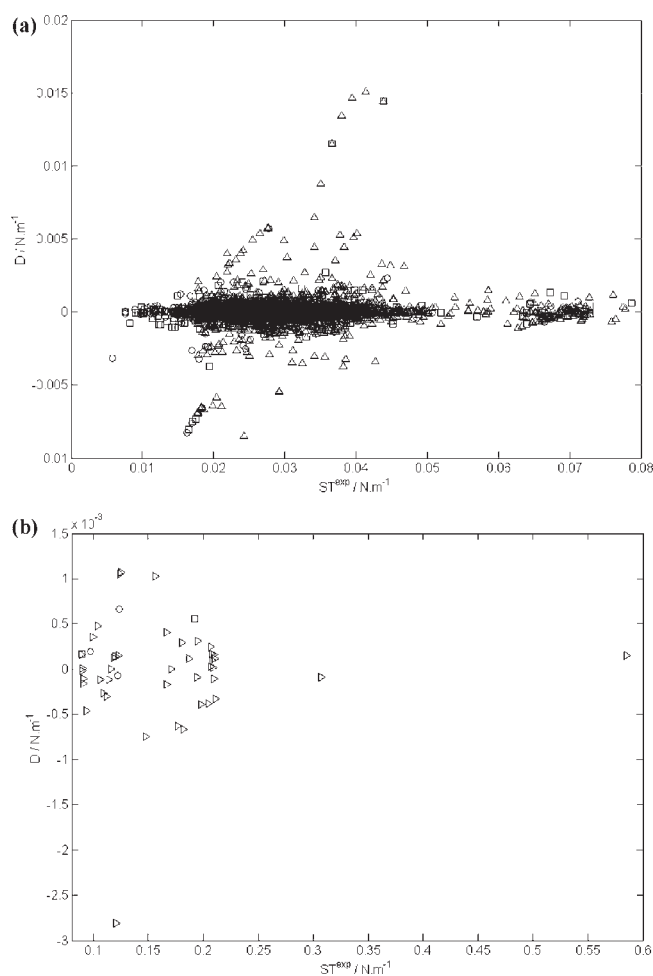
Figure 2. (a) Comparison between the represented/predicted results of the developed model and experimental values[30] of surface tensions, which are lower than those of water. (ST: surface tension of chemical compounds, N/m; D: deviation from experimental values = ($ST^{rep/pred}$ − ST^{exp}); rep: representation; pred: prediction; exp: experimental; Δ: training set, □: validation set, ○: test set.) (b) Comparison between the represented/predicted results of the developed model and experimental values[30] of surface tensions, which are higher than those of water. (ST: surface tension of chemical compounds, N/m; D: deviation from experimental values = ($ST^{rep/pred}$ − ST^{exp}); rep: representation; pred: prediction; exp: experimental; Δ: training set, □: validation set, ○: test set.)

which is capable of predicting the target values as accurately as possible. This step is repeated until the best ANN is obtained. Generally and especially in three-layer FFANNs, it is more efficient that the number of neurons in the hidden layer is optimized according to the accuracy of the obtained FFANN.[29−66]

3. RESULTS AND DISCUSSION

An optimized FFANN was obtained using the previously described procedure for representation/prediction of the surface tensions of 752 compounds at various temperatures and atmospheric pressure. For this purpose, several 3FFANNs modules were generated assuming numbers 1 through 50 for n (number of neurons in hidden layer) using the previously described procedure. The most accurate results were observed at $n = 10$. It should

Table 2. Statistical Parameters of the Presented Model[a]

statistical parameter	value
Training Set	The Model
R^2	0.997
Absolute average deviation [b]	1.7 %
Standard deviation error	0.02
Mean square error	0.0
N [c]	3738
Validation Set	
R^2	0.992
Absolute average deviation	1.7 %
Standard deviation error	0.01
Mean square error	0.0
N	467
Test Set	
R^2	0.995
Absolute average deviation	1.7 %
Standard deviation error	0.01
Mean square error	0.0
N	467
Training + Validation + Test Set	
R^2	0.997
Absolute average deviation	1.7 %
Standard deviation error	0.02
Mean square error	0.0
N	4672

[a] R^2: Squared correlation coefficient. [b] AAD % = $100/N\sum_i^N(|\text{Rep}(i)/\text{Pred}(i) - \text{Exp}(i)/\text{Exp}(i)|)$. [c] Number of data points.

be noted that this value is not the global value, because the optimization method used to train the ANN has great effects on the obtained value.[29−61] Therefore, the developed three-layer FFANN has the structure of 152-10-1 (temperature +151 chemical groups are regarded as the inputs of the algorithm). For every molecule, only few chemical groups are present and consequently used. To the best of our knowledge, there has been no temperature-dependent ANN-GC model for evaluation of physical properties of pure compounds in literature.

The *mat* file (MATLAB file format) of the obtained ANN containing all the parameters of the model (weight matrices and bias vectors) and the instruction for running the program is freely available upon request to the authors. The represented/predicted surface tensions are shown in Figure 2a,b in comparison with the experimental values.[30] We have splitted the data into two categories: 1. The surface tensions form the lowest value to that of 0.08 N/m, which is the surface tension of water; 2. Higher surface tension values than that of water. This splitting leads to show the capability of the proposed model for representation/ prediction of the surface tensions of the compounds with very low surface tension values at atmospheric pressure and different temperatures, which have been under debate regarding the to date presented methods.[1−26] The statistical results obtained by the ANN-GC model are reported in Table 2. As can be seen, the squared correlation coefficients, absolute average deviations, and standard deviation errors of the model over the "Training" set, the "Validation (Optimization)" set, the "Test (Prediction)" set, and the main data set are 0.997, 0.992, 0.995, and 0.997; 1.7 %,

Table 3. Average Absolute Deviations of the Obtained Results from Experimental Values[30] of Various Chemical Families

no.	family	AAD %
1	1-alkenes	1.0
2	2,3,4-alkenes	0.5
3	acetates	1.9
4	aldehydes	0.7
5	aliphatic ethers	2.3
6	alkylcyclohexanes	0.8
7	alkylcyclopentanes	1.7
8	alkynes	0.8
9	anhydrides	0.2
10	aromatic alcohols	1.8
11	aromatic amines	0.9
12	aromatic carboxylic acids	1.4
13	aromatic chlorides	1.6
14	aromatic esters	1.2
15	C, H, Br compounds	0.5
16	C, H, F compounds	1.7
17	C, H, I compounds	0.5
18	C, H, multihalogen compounds	1.0
19	C, H, NO_2 compounds	0.9
20	C1/C2 aliphatic chlorides	0.9
21	C3 and higher aliphatic chlorides	1.0
22	cycloaliphatic alcohols	2.5
23	cycloalkanes	2.4
24	cycloalkenes	1.0
25	dialkenes	0.4
26	dimethylalkanes	1.5
27	diphenyl/polyaromatics	3.9
28	elements	5.9
29	epoxides	0.4
30	ethyl and higher alkenes	1.5
31	formates	1.2
32	inorganic acids	1.7
33	inorganic gases	2.8
34	inorganic halides	9.2
35	ketones	1.0
36	mercaptans	1.1
37	methylalkanes	3.1
38	methylalkenes	1.3
39	multiring cycloalkanes	3.9
40	N-alcohols	1.0
41	N-aliphatic acids	0.9
42	N-aliphatic primary amines	1.6
43	N-alkanes	2.3
44	N-alkylbenzenes	0.8
45	naphthalenes	2.0
46	nitriles	0.6
47	nitroamines	7.7
48	organic salts	0.5
49	organic/inorganic compounds	0.2
50	other aliphatic acids	1.5
51	other aliphatic alcohols	1.3
52	other aliphatic amines	1.0

Table 3. Continued

no.	family	AAD %
53	other alkanes	2.6
54	other alkylbenzenes	1.6
55	other amines, imines	0.3
56	other condensed rings	1.8
57	other ethers/diethers	1.1
58	other hydrocarbon rings	1.8
59	other inorganics	2.2
60	other monoaromatics	1.2
61	other polyfunctional C, H, O	3.7
62	other polyfunctional organics	0.3
63	other saturated aliphatic esters	1.3
64	polyfunctional acids	0.3
65	polyfunctional amides/amines	0.3
66	polyfunctional C, H, N, halide, (O)	0.8
67	polyfunctional C, H, O, halide	1.3
68	polyfunctional C, H, O, N	0.6
69	polyfunctional C, H, O, S	0.3
70	polyfunctional esters	1.2
71	polyfunctional nitriles	0.3
72	polyols	0.9
73	propionates and butyrates	1.3
74	silanes/siloxanes	6.2
75	sodium salts	0.3
76	sulfides/thiophenes	0.8
77	terpenes	0.5
78	unsaturated aliphatic esters	1.3

1.7 %, 1.7 %, and 1.7 %; and 0.02, 0.01, 0.01, and 0.02, respectively. Furthermore, the absolute average deviations of the results from experimental values[30] for each 78 chemical families are reported in Table 3. The results imply that the obtained ANN-GC model is accurate to represent/predict the surface tensions of the investigated pure compounds from various chemical families at different temperatures. For better illustration of the pure compounds investigated in this study, their chemical structures are sketched in the Supporting Information.

To omit from our discussion that there are 105 data points (from 4672 available data points) for which the presented model results lead to more than 10 % absolute deviations from experimental values,[30] would be an oversight. It seems that there is no relation between these compounds structures to show some weaknesses in representing/predicting of the surface tension values of related chemical families. Moreover, the highest deviations (more than 40 %) are from the same data source (refer to the Supporting Information). Therefore, it is probable that the surface tension values for these compounds are not accurate or may be somehow erroneous because of the existed difficulties and possible errors in experimental measurements. It should be noted that the proposed method has been developed based of all of the available data including the aforementioned 105 data points.

4. CONCLUSION

In this work, a group contribution-based model was presented for representation/prediction of the surface tensions of pure compounds (containing organic and inorganic liquids) at different temperatures (temperature dependent ANN-GC model)

and atmospheric pressure. The model is the result of combination of feed forward neural networks and group contributions. The required parameters of the model are temperature and the numbers of occurrences of 151 functional groups in each investigated molecule. It should be noted that most of these functional groups are not simultaneously available in a particular molecule. Therefore, computation of the required parameters from chemical structure of any molecule is simple. For developing the model, the experimental surface tension values from the largest available data set[30] containing 752 pure compounds from various chemical families and at different temperatures (totally 4672) data points were applied. As a consequence, a reliable, comprehensive, and predictive tool was developed to represent/predict the surface tensions of many of pure compounds, which are especially applied in chemical and petroleum industries. However, one point should not be omitted form our conclusion: The model has a wide range of applicability but the prediction capability of the model is restricted to the compounds, which are similar to those ones applied to develop the model. Application of the model for the totally different compounds than the investigated ones is not recommended although it may be used for a rough estimation of the surface tensions of these kinds of compounds.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. The number of occurrences of the 151 functional groups in all of 752 pure compounds in the main data set, the distributions of the data in three data sets, the obtained results, and sources of experimental data in the first excel file, and the schematic structures of all of the investigated compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
*E-mail: amir-hossein.mohammadi@mines-paristech.fr. Tel.: + (33) 1 64 69 49 70. Fax: + (33) 1 64 69 49 68.

■ ACKNOWLEDGMENT

■ REFERENCES

(1) Balasubrahmanyam, S. N. Einstein, 'parachor' and molecular volume: Some history and a suggestion. *Curr. Sci. India* **2008**, *94*, 1650–1658.

(2) Danesh, A. *PVT and Phase Behaviour of Petroleum Reservoir Fluids*; Elsevier: Netherlands, 1998.

(3) Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *Properties of Gases and Liquids*, 5th ed.; McGraw-Hill: New York, 2001.

(4) Escobedo, J.; Mansoori, A. Surface tension prediction for pure fluids. *AIChE J.* **1996**, *42*, 1425–1433.

(5) Macleod, D. B. Relation between surface tension and density. *Trans. Faraday Soc.* **1923**, *19*, 38–42.

(6) Sugden, S. A Relation between surface tension, density, and chemical composition. *J. Am. Chem. Soc.* **1924**, *125*, 1177–1189.

(7) Sugden, S. The Parachor and Valency; G. Routledge Sons Ltd.: London. 1930.

(8) Bayliss, N. S. Atomic radii from parachor data and from electron diffraction data. *J. Am. Chem. Soc.* **1937**, *59*, 444–447.

(9) Schechter, D. S.; Guo, B. Parachors based on modern physics and their uses in IFT prediction of reservoir fluids. *SPE 30785* Proc. of Ann. Conf., Dallas, USA, 1995.

(10) Baker, O.; Swerdloff, W. Calculations of surface tension-3: Calculations of surface tension parachor values. *Oil Gas J.* 1955, 43, 141.

(11) Fanchi, J. R. Calculation of parachors for compositional simulation, an update. *SPE Res. Eng.* Aug. 1990, 5, 433-436.

(12) Quayle, R. The parachor of organic compounds. *Chem. Rev.* 1953, 53439–53589.

(13) Weinaug, C. F.; Kalz, D. L. Surface tensions of methane-propane mixtures. *Ind. Eng. Chem.* 1943, 35, 239–246.

(14) Lee, S.-T.; Chien, M. C. H. A new multicomponent surface tension correlation based on scaling theory. SPE/DOE Symp. on EOR, SPE/DOE 12643, Tulsa, OK (Apr. 15−18, 1984).

(15) Hugill, J. A.; van Welsenes, A. J. Surface tension: A simple correlation for natural gas + condensate systems. *Fluid Phase Equilib.* 1986, 29, 383–390.

(16) Gasem, K. A. M.; Dulcamara, P. B.; Dickson, B. K.; Robinson, R. L., Jr. Test of prediction methods for interfacial tensions of CO, and ethane in hydrocarbon solvents. *Fluid Phase Equilib.* 1989, 53, 39–50.

(17) Fanchi, J. R. Calculation of parachors for compositional simulation. *JPT* 1985, 85, 2049–2050.

(18) Ali, J. K. Prediction of parachors of petroleum cuts and pseudo-components. *Fluid Phase Equilib.* 1994, 95, 383–398.

(19) Brock, J. R.; Bird, R. B. Surface tension and the principle of corresponding states. *AIChE. J.* 1955, 1, 174–177.

(20) Guggenheim, E. A. The principle of corresponding states. *J. Chem. Phys.* 1945, 13, 253–262.

(21) Curl, R. F.; Pitzer, K. Volumetric and thermodynamic properties of fluids—enthalpy, free energy, and entropy. *Ind. Eng. Chem.* 1958, 50, 265–274.

(22) Pitzer, K. S.: *Thermodynamics*, 3d ed., New York, McGraw-Hill, 1995.

(23) Zuo, Y.-X.; Stenby, E. H. Corresponding-states and parachor models for the calculation of interfacial tensions. *Can. J. Chem. Eng.* 1997, 75, 1130–1137.

(24) Rice, P.; Teja, A. S. A generalized corresponding-states method for the prediction of surface tension of pure liquids and liquid mixtures. *J. Colloid Interf. Sci.* 1982, 86, 158–163.

(25) Sastri, S. R. S.; Rao, K. K. A simple method to predict surface tension of organic liquids. *Chem. Eng. J.* 1995, 59, 181–186.

(26) Freitas, A. A.; Quina, F. H.; Carroll, F. A. Estimation of water-organic interfacial tensions. A linear free energy relationship analysis of interfacial adhesion. *J. Phys. Chem. B* 1997, 101, 7488–7493.

(27) Kavun, S. M.; Chalykh, A. E.; Palyulin, V. A. Prediction of surface tension of organic liquids. *Colloid. J. Russ. Acad.* 1995, 57, 767–771.

(28) Riedel, L. Eine neue universelle Dampfdruckformel. Untersuchungen über eine Erweiterung des Theorems der übereinstimmenden Zustände. Teil I. *Chem. Ing. Technol.* 1954, 26, 83–89.

(29) Gharagheizi, F.; Sattari, M. Prediction of triple-point temperature of pure components using their chemical structures. *Ind. Eng. Chem. Res.* 2010, 49, 929–932.

(30) Project 801, Evaluated Process Design Data, Public Release Documentation, Design; Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE), 2006.

(31) Gharagheizi, F.; Abbasi, R.; Tirandazi, B. Prediction of Henry's law constant of organic compounds in water from a new group-contribution-based model. *Ind. Eng. Chem. Res.* doi: 10.1021/ie101532e.

(32) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Artificial neural network modeling of solubilities of 21 commonly used industrial solid compounds in supercritical carbon dioxide. *Ind. Eng. Chem. Res.* 2011, 50, 221–226.

(33) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Representation/prediction of solubilities of pure compounds in water using artificial neural network - group contribution method. *J. Chem. Eng. Data*, Article ASAP, doi: 10.1021/je100619p, 2011.

(34) Chouai, A.; Laugier, S.; Richon, D. Modeling of thermodynamic properties using neural networks: Application to refrigerants. *Fluid Phase Equilib.* 2002, 199, 53–62.

(35) Piazza, L.; Scalabrin, G.; Marchi, P.; Richon, D. Enhancement of the extended corresponding states techniques for thermodynamic modelling. I. Pure fluids. *Int. J. Refrig.* 2006, 29, 1182–1194.

(36) Scalabrin, G.; Marchi, P.; Bettio, L.; Richon, D. Enhancement of the extended corresponding states techniques for thermodynamic modelling. II. Mixtures. *Int. J. Refrig.* 2006, 29, 1195–1207.

(37) Chapoy, A.; Mohammadi, A. H.; Richon, D. Predicting the hydrate stability zones of natural gases using Artificial Neural Networks. *Oil Gas Sci. Technol. − Rev. IFP* 2007, 62, 701–706.

(38) Mohammadi, A. H.; Richon, D. Hydrate phase equilibria for hydrogen + water and hydrogen + tetrahydrofuran + water systems: Predictions of dissociation conditions using an artificial neural network algorithm. *Chem. Eng. Sci.* 2010, 65, 3352–3355.

(39) Mohammadi, A. H.; Richon, D. Estimating sulfur content of hydrogen sulfide at elevated temperatures and pressures using an artificial neural network algorithm. *Ind. Eng. Chem. Res.* 2008, 47, 8499–8504.

(40) Mohammadi, A. H.; Richon, D. A Mathematical model based on artificial neural network technique for estimating liquid water-hydrate equilibrium of water - hydrocarbon System. *Ind. Eng. Chem. Res.* 2008, 47, 4966–4970.

(41) Mohammadi, A. H.; Afzal, W.; Richon, D. Determination of critical properties and acentric factors of petroleum fractions using artificial neural networks. *Ind. Eng. Chem. Res.* 2008, 47, 3225–3232.

(42) Mohammadi, A. H.; Richon, D. Use of artificial neural networks for estimating water content of natural gases. *Ind. Eng. Chem. Res.* 2007, 46, 1431–1438.

(43) Mohammadi, A. H.; Martínez-López, J. F.; Richon, D. Determining phase diagrams of tetrahydrofuran+methane, carbon dioxide or nitrogen clathrate hydrates using an artificial neural network algorithm. *Chem. Eng. Sci.* 2010, 65, 6059–6063.

(44) Mehrpooya, M.; Mohammadi, A. H.; Richon, D. Extension of an artificial neural network algorithm for estimating sulfur content of sour gases at elevated temperatures and pressures. *Ind. Eng. Chem. Res.* 2010, 49, 439–442.

(45) Mohammadi, A. H.; Belandria, V.; Richon, D. Use of an artificial neural network algorithm to predict hydrate dissociation conditions for hydrogen + water and hydrogen + tetra-n-butyl ammonium bromide + water systems. *Chem. Eng. Sci.* 2010, 65, 4302–4305.

(46) Gharagheizi, F. A new group contribution-based method for estimation of lower flammability limit of pure compounds. *J. Hazard. Mater.* 2009, 170, 595–604.

(47) Gharagheizi, F. New Neural Network Group Contribution model for estimation of lower flammability limit temperature of pure compounds. *Ind. Eng. Chem. Res.* 2009, 48, 7406–7416.

(48) Gharagheizi, F.; Sattari, M. Estimation of molecular diffusivity of pure chemicals in water: A quantitative structure-property relationship study. *SAR QSAR Environ. Res.* 2009, 20, 267–285.

(49) Gharagheizi, F. Prediction of standard enthalpy of formation of pure compounds using molecular structure. *Aust. J. Chem.* 2009, 62, 376–381.

(50) Gharagheizi, F.; Tirandazi, B.; Barzin, R. Estimation of aniline point temperature of pure hydrocarbons: A Quantitative Structure-Relationship approach. *Ind. Eng. Chem. Res.* 2009, 48, 1678–1682.

(51) Gharagheizi, F.; Mehrpooya, M. Prediction of some important physical properties of sulfur compounds using QSPR models. *Mol. Divers.* 2008, 12, 143–155.

(52) Sattari, M.; Gharagheizi, F. Prediction of Molecular diffusivity of pure components into air: A QSPR approach. *Chemosphere* 2008, 72, 1298–1302.

(53) Gharagheizi, F.; Alamdari, R. F.; Angaji, M. T. A new neural network-group contribution method for estimation of flash point. *Energy Fuel.* 2008, 22, 1628–1635.

(54) Gharagheizi, F.; Fazeli, A. Prediction of Watson characterization factor of hydrocarbon compounds from their molecular properties. *QSAR Comb. Sci.* 2008, 27, 758–767.

(55) Gharagheizi, F. A New Neural Network Quantitative Structure-Property Relationship for prediction of θ (Lower Critical Solution Temperature) of polymer solutions. *e-Polymer* 2007, 114, 1–5.

(56) Gharagheizi, F. QSPR studies for solubility parameter by means of genetic algorithm-based multivariate linear regression and generalized regression neural network. *QSAR Comb. Sci.* **2008**, *27*, 165–170.

(57) Gharagheizi, F. A chemical structure-based model for estimation of upper flammability limit of pure compounds. *Energ. Fuel.* **2010**, *27*, 3867–3871.

(58) Vatani, A.; Mehrpooya, M.; Gharagheizi, F. Prediction of standard enthalpy of formation by a QSPR Model. *Int. J. Mol. Sci.* **2007**, *8*, 407–432.

(59) Mehrpooya, M.; Gharagheizi, F. A molecular approach for prediction of sulfur compounds solubility parameters, phosphorus sulfur and silicon and related elements. *Phosphorus Sulfur* **2010**, *185*, 204–210.

(60) Gharagheizi, F.; Alamdari, R. F. A molecular-based model for prediction of solubility of c60 fullerene in various solvents. *Fullerenes, Nanotubes, Carbon, Nanostruct.* **2008**, *16*, 40–57.

(61) Eslamimanesh, A.; Gharagheizi, F.; Mohammadi, A. H.; Richon, D. Artificial neural network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. *Chem. Eng. Sci.* **2011**, accepted.

(62) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Determination of parachor of various compounds using artificial neural network-group contribution approach. *Ind. Eng. Chem. Res.* **2011**, accepted.

(63) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Determination of critical properties and acentric factors of pure compounds using artificial neural network-group contribution algorithm. *J. Chem. Eng. Data* **2011**, doi: 10.1021/je200019g.

(64) Kalogirou, S. A. Artificial Neural Networks in renewable energy systems applications: A review. *Renew. Sust. Energ. Rev.* **2001**, *5*, 373–401.

(65) Hagan, M.; Demuth, H. B.; Beale, M. H. Neural Network Design; International Thomson publishing, 2002.

(66) Gharagheizi, F. QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN. *Comput. Mater. Sci.* **2007**, *40*, 159.