JOURNAL OF CHEMICAL & ENGINEERING DATA

ARTICLE

pubs.acs.org/jced

Determination of Critical Properties and Acentric Factors of Pure Compounds Using the Artificial Neural Network Group Contribution Algorithm

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

[†]Saman Energy Giti Co., Postal Code 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: In this article, artificial neural network group contribution (ANN-GC) method is applied to calculate and estimate critical properties including the critical pressure, temperature, and volume and acentric factors of pure compounds. About 1700 chemical compounds from various chemical families have been investigated to propose a comprehensive and predictive model. Using this dedicated model, we obtain satisfactory results quantified by the following absolute average deviations of the calculated and estimated properties from existing experimental values: 1.1 % for critical pressure, 0.9 % for critical temperature, 1.4 % for critical volume, and 3.7 % for acentric factor.

1. INTRODUCTION

Physical properties of chemical compounds, including the critical temperature (T_c), pressure (P_c), and volume (V_c), and also the acentric factor (ω) are of great interest for researchers and engineers in the past and present years mainly for designing and modeling chemical processes. The applications of such properties can be summarized as follows:

1 Corresponding states theory;[1]
2 Determination of the equation-of-state parameters for modeling the phase behavior of the systems encountered in the processes, for example, oil refineries, petrochemical companies, pharmaceutical, and food engineering processes, and so forth;[1,2]
3 Modeling the fluid transport systems in industrial projects, for example, oil and gas production, petroleum fluid transportation;[1,2] and,
4 Determination of thermal and transport properties such as enthalpies, heats of vaporization, viscosities, and interfacial tension.[3]

However, experimental determination of these property values is time-consuming, costly, and sometimes a challenge since some of high molecular weight or larger compounds may chemically degrade before they reach critical conditions.[1,3−5] Consequently, representation and prediction methods have been widely presented in the literature for providing the property values when experimental ones are not available.

Most of the calculation and estimation methods for critical properties and acentric factors are based on group contributions (GC), that is, the properties of a molecule are normally established from contributions from its elements.[1] The conceptual basis is that the intermolecular forces that determine the constants of interest depend mostly on the bonds between the atoms of the molecules.[1] Therefore, each molecule is treated as to be composed of fundamental groups, each one giving a contribution to the property of interest, which is calculated by adding together the contributions of each group.[5]

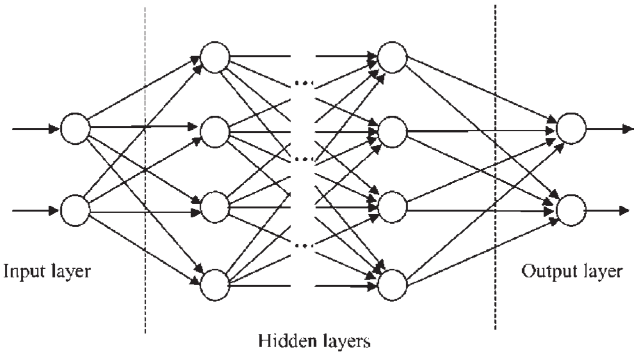


Figure 1. Schematic structure of the FFANN used in this study.[65]

So far, many group-contribution methods have been proposed such as the methods by Kudchadker and Zwolinski,[6] Lydersen,[7] Joback and Reid,[8] Marrero and Gani,[9] Kudchadker et al.,[10] Thodos,[11−14] Ambrose and Ghiasse,[15] Wilson and Jasperson,[16] Marrero-Morejón and Pardillo-Fontdevila,[17] Lydersen and Tsochev,[18] and Dalmazzone et al.[19] The method proposed by Thodos[11,12] is based on van der Waals interaction and repulsion parameters of paraffins, which have later extended to other homologous series.[3,13,14] One of the first successful group

Received: January 7, 2011
Accepted: February 24, 2011
Published: April 04, 2011

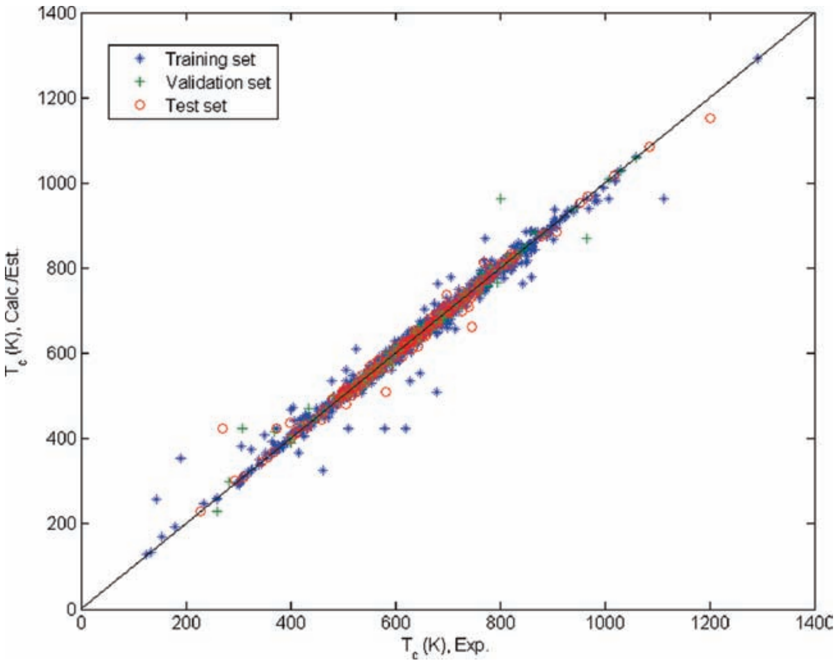


Figure 2. Comparison between the calculated and estimated results of the first model and experimental values[32] of critical temperatures.

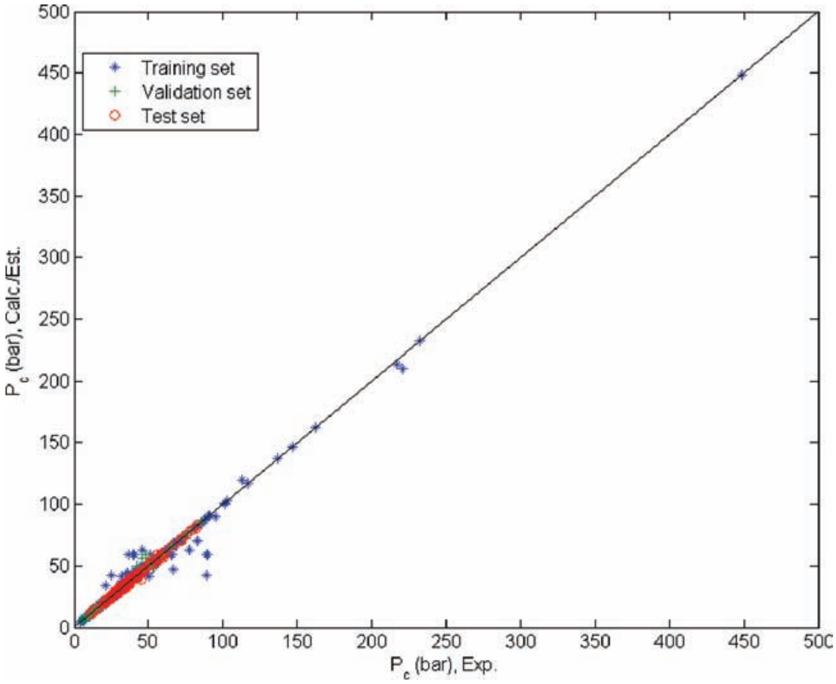


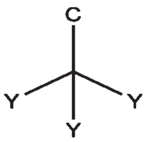
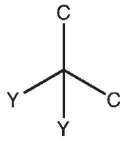
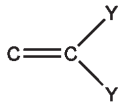
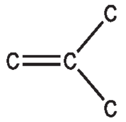
Figure 3. Comparison between the calculated and estimated results of the first model and experimental values[32] of critical pressures.

contribution methods to estimate critical properties has been developed by Lydersen[7] in 1955.[1] As, the capabilities of the mathematical correlations and computer techniques have been improved, more accurate optimized values of the models have been obtained. Examples are the methods of Joback,[20] Constantinou and Gani,[21] Wilson and Jasperson,[16] and Marrero-Morejon and Pardillo-Fontdevila.[17]

Joback[20] reevaluated Lydersen's group contribution scheme,[7] added several new functional groups, determined new contribution values, and used normal boiling point temperature and number of atoms for estimation of the critical properties.[1] There are other similar methods presented by Somayajulu,[22] who has also utilized the boiling point to estimate critical properties.[3]

Constantinou and Gani[21] developed an advanced group contribution method based on the universal functional activity coefficient (UNIFAC) groups. They added second-order group contributions to account for configuration such as isomers, multiple groups located close together, resonance structures,

Table 1. Functional Groups Used to Develop the Model for Calculation/Estimation of Properties and Acentric Factors of 2-Methyl-1-pentene

GC ID[a]	Definition of GC	Number of GCs
001		2
002		2
010		1
012		1
108	H attached to C^1(sp3) / C^0(sp2)[b]	2

[a] Group contribution identification (for observing complete list of the group contributions, see the Supporting Information files). [b] The superscript represents the formal oxidation number.

Table 2. Statistical Parameters of the Presented Models for Critical Temperature

statistical parameter	value	value
Training Set	First Model	Second Model
R^{2a}	0.983	0.994
absolute average deviation[b]	1.5 %	0.9 %
standard deviation error	131.6	134.0
mean square error	295.3	107.2
N^c	1373	1359
Validation Set		
R^2	0.977	0.996
absolute average deviation	1.6 %	0.8 %
standard deviation error	130.81	117.6
mean square error	389.3	60.5
N	171	169
Test Set		
R^2	0.984	0.992
absolute average deviation	1.8 %	1.0 %
standard deviation error	143.65	128.9
mean square error	355	132.5
N	171	169
Training + Validation + Test Set		
R^2	0.983	0.994
absolute average deviation	1.5 %	0.9 %
standard deviation error	132.74	131.9
mean square error	310.6	105.1
N	1715	1697

[a] Squared correlation coefficient. [b] % AAD = $(100/N)\Sigma_i^N((|\mathrm{Calc.}(i)/\mathrm{Est.}(i) - \mathrm{Exp.}(i)|)/\mathrm{Exp.}(i))$. [c] Number of data points.

and so forth.[1] Their method has been widely used in chemical engineering problems.[1]

Several authors (Kreglewski and Zwolinski,[23] Gasem,[24] Gasem et al.,[25] Tsonopolous and Tan,[26] Magoulas and Tassios,[27] Teja and Smith,[28] Marano and Holder,[29] and Morgan and Kobayashi[30]) have developed correlations for critical properties based on the number of carbon atoms. For instance, Gasem and co-workers[25] developed the asymptotic behavior correlation (ABC) model framework to predict the physical properties of n-paraffins applying number of carbon atoms and two scaling coefficients. Correlations between critical pressure and temperature and molecular weight have been proposed by Kontogeorgis and Tassios.[31] Comparison between the aforementioned methods regarding the number of the investigated compounds, the average absolute deviations, and range of applicability are well reported in the *Properties of Gases and Liquids* by Poling et al.[1] and elsewhere.[3,5]

In spite of the fact that these calculation and estimation methods are simple to apply, they have significant drawbacks:

- Not all of groups encountered in various chemical compounds are investigated; that is, these methods do not cover a wide range of chemical compounds;
- The results for stereoisomers are not conclusive;[5]
- They are not adequately accurate.[1,3−5]

Recently, the quantitative structure−property relationship (QSPR) models have been applied to overcome the following drawbacks.[3,5] A QSPR model consists of a correlation between the property of interest and a variety of molecular features (named descriptors) that range from structural and topological indices to electronic and quanto-chemical properties.[3,5] Godavarthy et al.[3] proposed a QSPR model for the calculation and estimation of the critical properties. They obtained reliable results with average absolute deviation of 3.7 % for T_c, 0.49 % for P_c, and 5.2 % for V_c regarding 1230 chemical compounds. Another QSPR approach was done by Sola and co-workers,[5] who have studied 155 chemical compounds for the representation of critical pressure and temperature. They have reported the root-mean-square errors of (12.6 and 9.66) K regarding the "Training" and "Validation" sets, respectively for critical temperature results and those of (0.25 and 0.28) MPa for critical pressure results.

Even with the use of the two mentioned QSPR methods, the average deviations are still high compared to the experimental values, and the models are not as comprehensive and predictive as expected. In this work, we present a new approach based on the artificial neural network group contribution method (ANN-GC) for the calculation and estimation of critical properties and acentric factors of around 1700 chemical compounds.

2. MATERIALS AND METHODS

2.1. Materials. The DIPPR 801[32] database has special applications in developing new methods for the representation and prediction of physical properties because it contains a large number of pure compounds with their evaluated physical properties. This database has been used here to provide a data set for calculation/estimation of critical properties and acentric factors. About 1700 pure compounds from 81 different chemical groups are investigated, and the related values of the critical temperature,

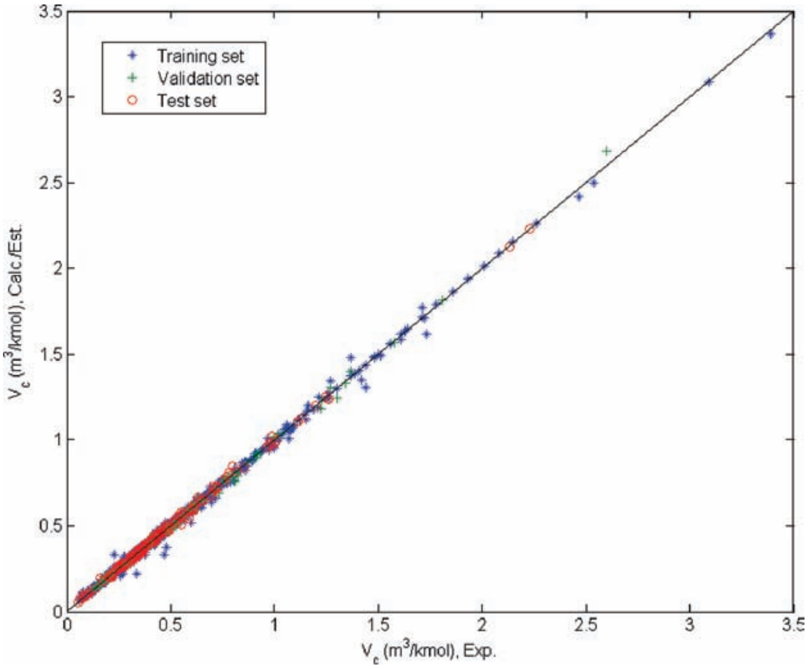


Figure 4. Comparison between the calculated and estimated results of the first model and experimental values[32] of critical volumes.

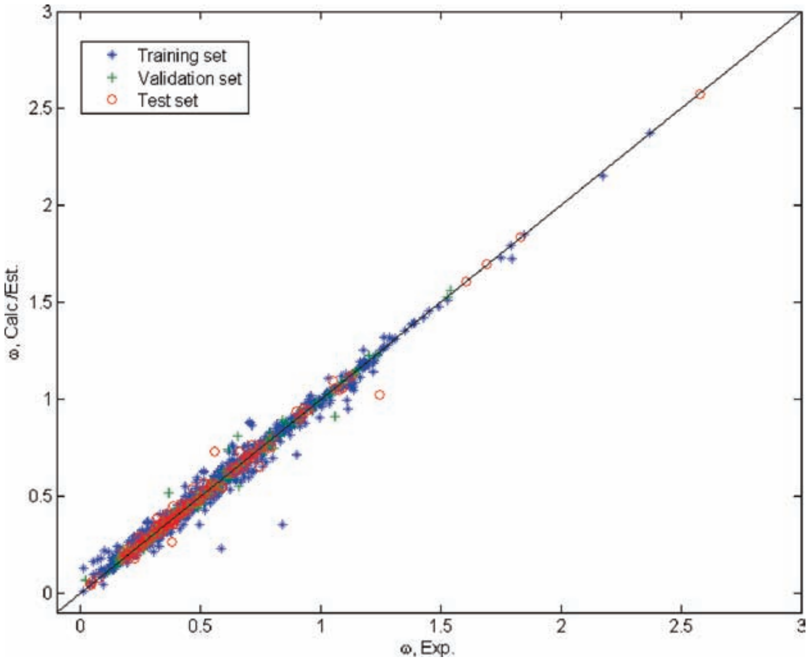


Figure 5. Comparison between the calculated and estimated results of the first model and experimental values[32] of acentric factors.

pressure, volume, and acentric factors are considered for the calculations. The names of the investigated compounds, formulas, and chemical families and the uncertainties of the experimental values are presented as Supporting Information.

2.2. Development of New Group Contributions. Having defined the database, the chemical structures of all of the studied compounds are analyzed using an algorithm comparing the chemical groups to define the most efficient contributions for evaluation of the desired property. Consequently, 149 functional groups have been found to be more efficient for the representation and prediction of the critical temperatures, 149 for the critical pressures, 166 for the critical volume, and 166 for acentric factor of the investigated pure compounds. The numbers of the group contributions are different due to the slightly different number of chemical compounds for the calculation and estimation of each property. This is mainly because of the different number of evaluated data points for each property in the DIPPR 801[32] data set. It should be noted that, for the representation and

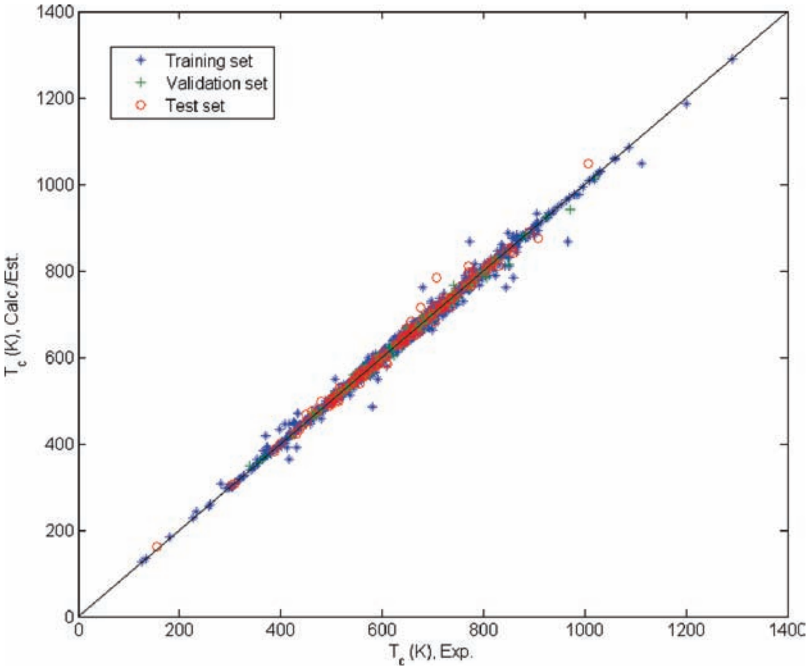


Figure 6. Comparison between the calculated and estimated results of the second model and experimental values[32] of critical temperatures.

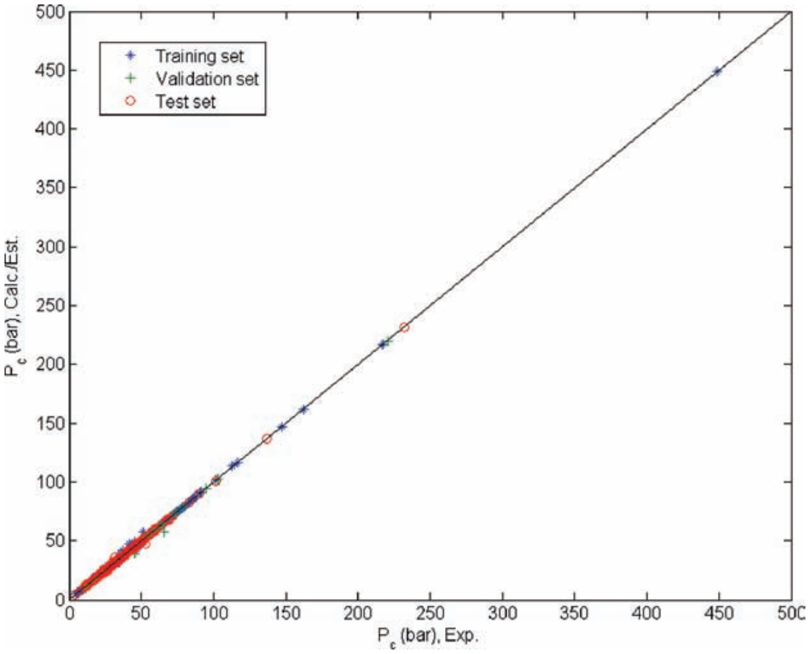


Figure 7. Comparison between the calculated and estimated results of the second model and experimental values[32] of critical pressures.

prediction of properties, the same chemical groups are selected. The functional groups used in this study are presented as Supporting Information. The tables of their numbers of occurrences in pure compounds used in this work have been also presented as Supporting Information. Table 1 shows an example of occurrences of the group contributions in 2-methyl-1-pentene. These chemical groups are used as the proposed model parameters.

2.3. Generation of ANN-GC. 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 the assumption of the existence of a multilinear relationship between these groups and the desired property (here the critical properties and acentric factor).[33] This technique is a similar method used in the most of classical group contribution methods.[34] Several calculations show that the application of the mentioned methodology for the current problem brings about poor results. Consequently, the nonlinear mathematical method of artificial neural network (ANN) is preferred and investigated.

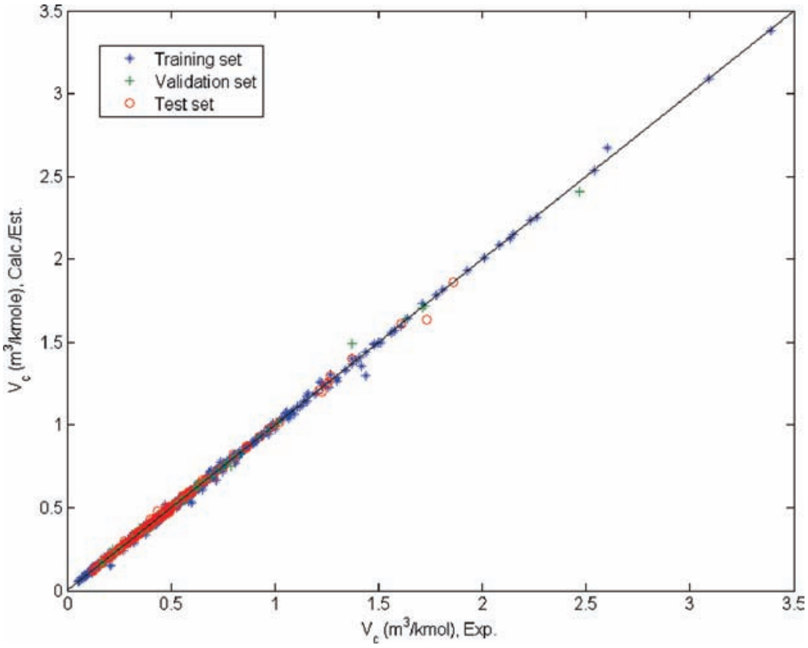


Figure 8. Comparison between the calculated and estimated results of the second model and experimental values[32] of critical volumes.

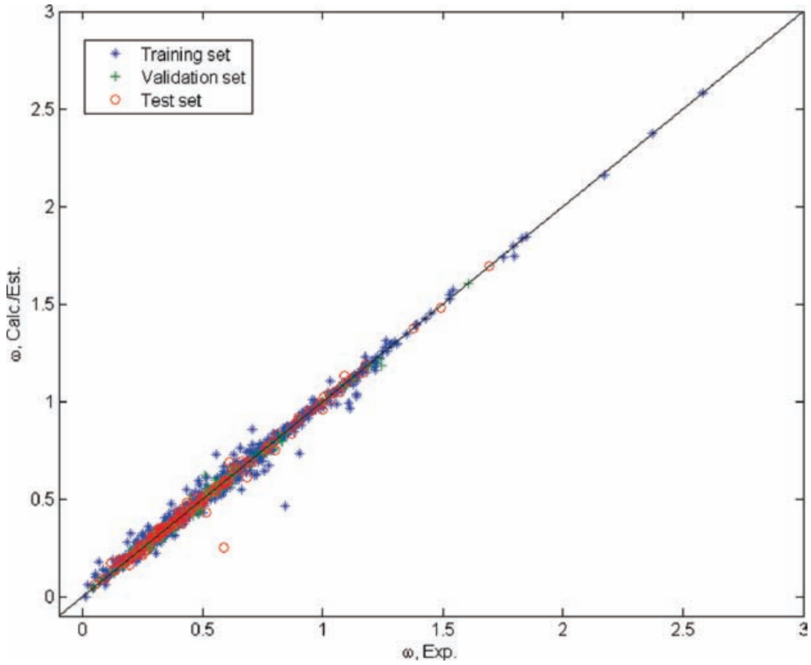


Figure 9. Comparison between the calculated and estimated results of the second model and experimental values[32] of acentric factors.

ANNs are extensively used in various scientific and engineering problems,[33−65] for example, calculations and estimations of physical and chemical properties of different pure compounds.[48−61,63,64] These capable mathematical tools are generally applied to study the complicated systems, e.g., phase behavior of fluid mixtures containing clathrate/semiclathrate hydrates.[41,43,45,47] The theoretical explanations about neural networks can be found elsewhere.[66] Using the ANN toolbox of the MATLAB software (Mathworks Inc.), a three-layer feed forward artificial neural network (FFANN) has been developed for the problem. The typical structure of a FFANN is schematically presented in Figure 1. The capabilities of this kind of ANNs have been demonstrated in previous works.[33−65]

All of the functional groups and also the property 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 desired properties for output parameters. Later, the database is divided into three subdata sets including the "Training" set, the "Validation"

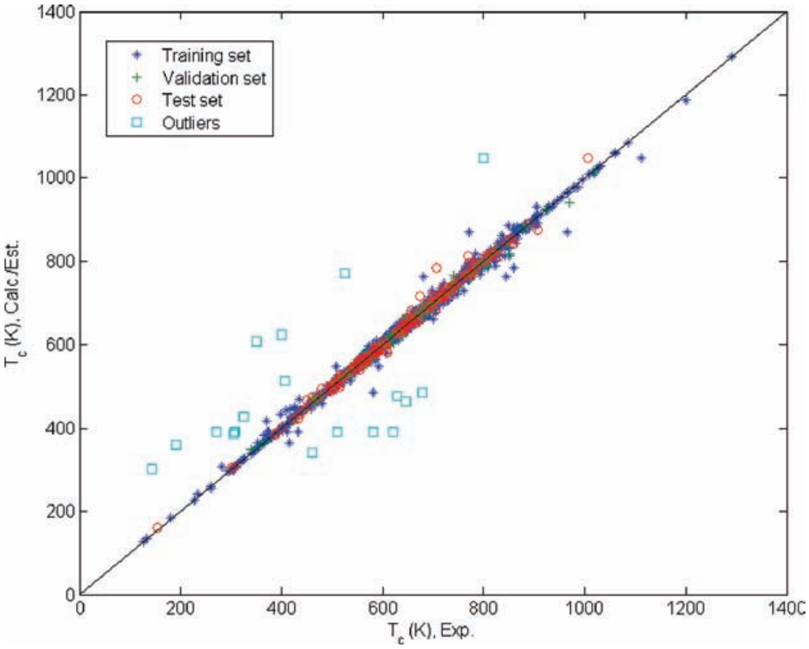


Figure 10. Second model predicted results of critical temperatures for outlier points (□) of the first model.

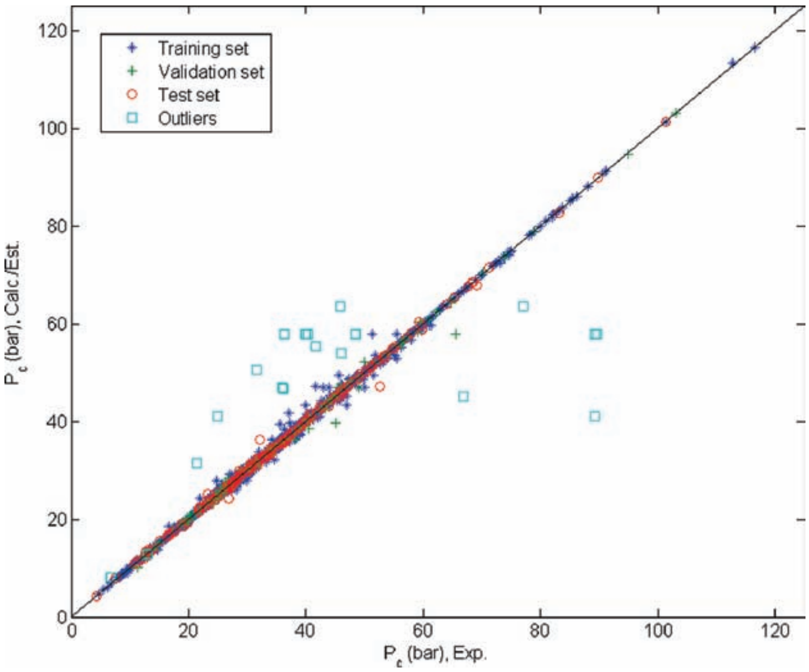


Figure 11. Second model predicted results of critical pressures for outlier points (□) of the first model.

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 1370 compounds), the "Validation" set (around 170 compounds), and the "Test" set (about 170 compounds). 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.[66,67]

As a matter of fact, generating an ANN model is determination of the weight matrices and bias vectors.[33−65] 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 (calculated and estimated properties) and the target values (experimental critical properties and acentric factors). This minimization is performed by Levenberg−Marquardt (LM)[66] optimization strategy. There

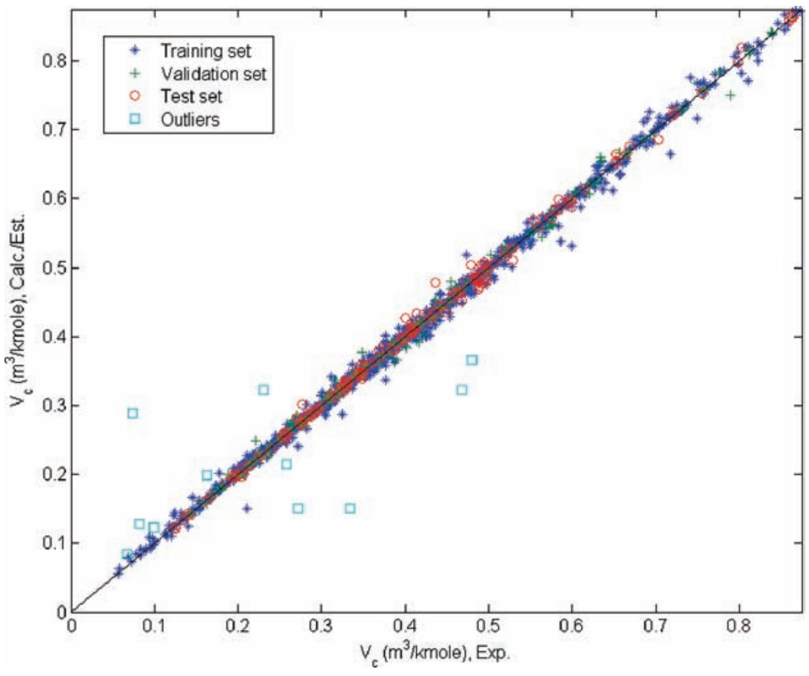


Figure 12. Second model predicted results of critical volumes for outlier points (□) of the first model.

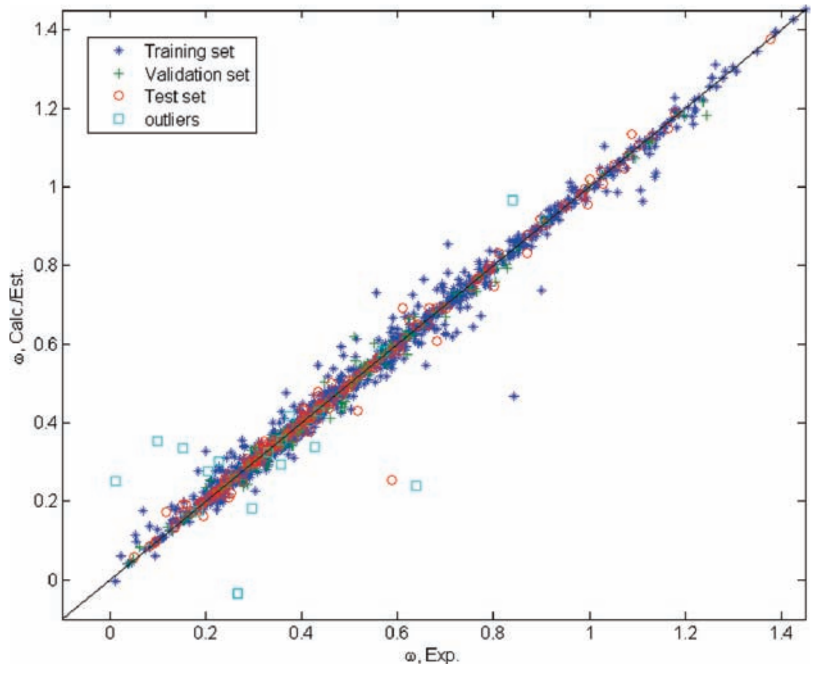


Figure 13. Second model predicted results of acentric factors for outlier points (□) of the first model.

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[66] is most-widely used algorithm for training due to being robust and accurate enough to deal with the considered system.[33−65]

In most cases, the number of neurons in the hidden layer (n) is fixed. Therefore, the main goal is to produce an ANN model, which is able to predict 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.[33−61]

3. RESULTS AND DISCUSSION

An optimized FFANN has been obtained using the aforementioned procedure for the calculation and estimation of the critical properties and acentric factors of studied compounds. For this

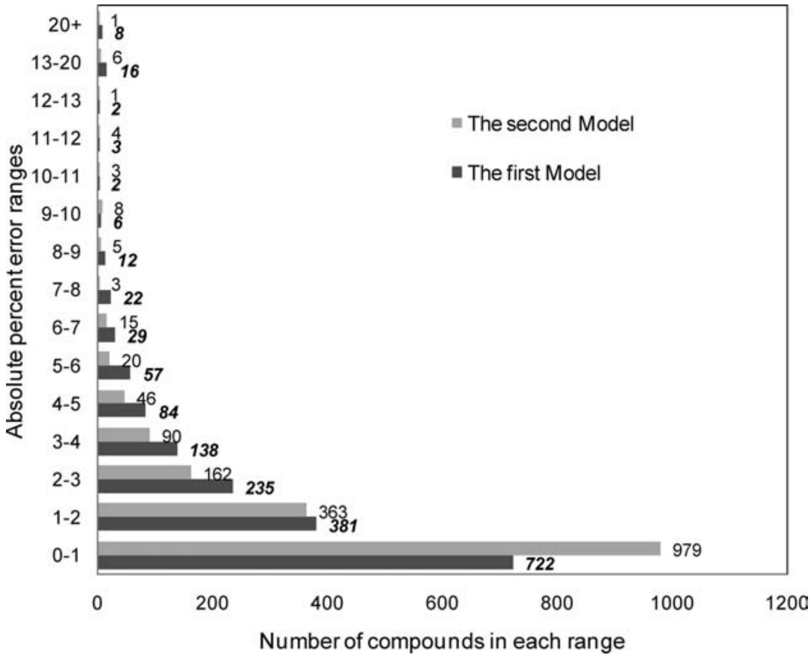


Figure 14. Deviation ranges of the results of the two developed models over all of the investigated compounds for critical temperatures.

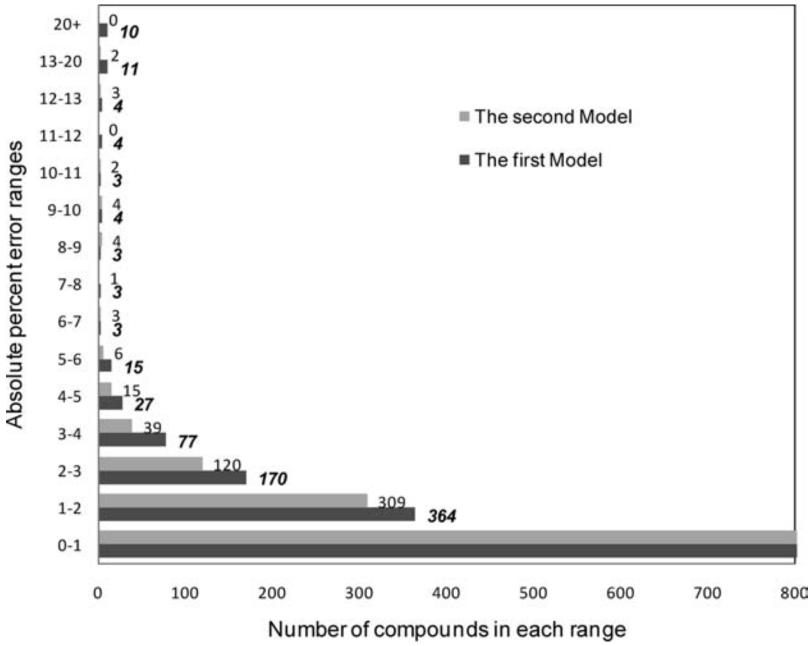


Figure 15. Deviation ranges of the results of the two developed models over all of the investigated compounds for critical pressures.

purpose, several 3FFANNs modules have been generated assuming numbers 1 through 50 for n (number of neurons in hidden layer) using the previously described procedure. The most accurate results without overfitting are observed for $n = 15$. It should 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.[32] Therefore, the developed three-layer FFANNs have the structure of 149-15-1, 149-15-1, 166-15-1, and 166-15-1 regarding the T_c, P_c, V_c, and ω.

The mat file (MATLAB file format) of the obtained ANNs containing all of the parameters of the model has been presented as Supporting Information. Moreover, the instruction for running the program (mat file) is presented in the Appendix. The calculated and estimated properties are shown in Figures 2 to 5 in comparison with the experimental values.[32] A more meticulous investigation of the results show that there are some compounds for which the presented model results lead to much higher absolute deviations from experimental values.[32] There are two

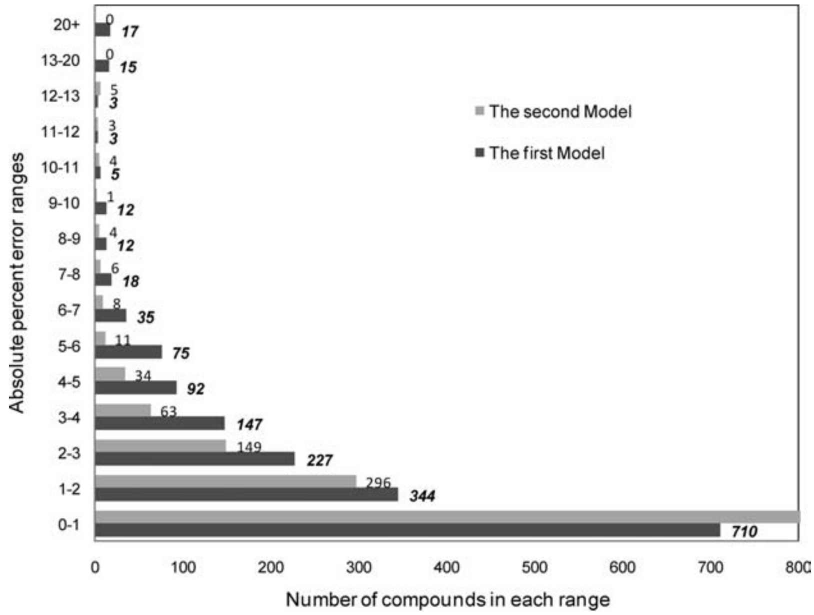


Figure 16. Deviation ranges of the results of the two developed models over all of the investigated compounds for critical volumes.

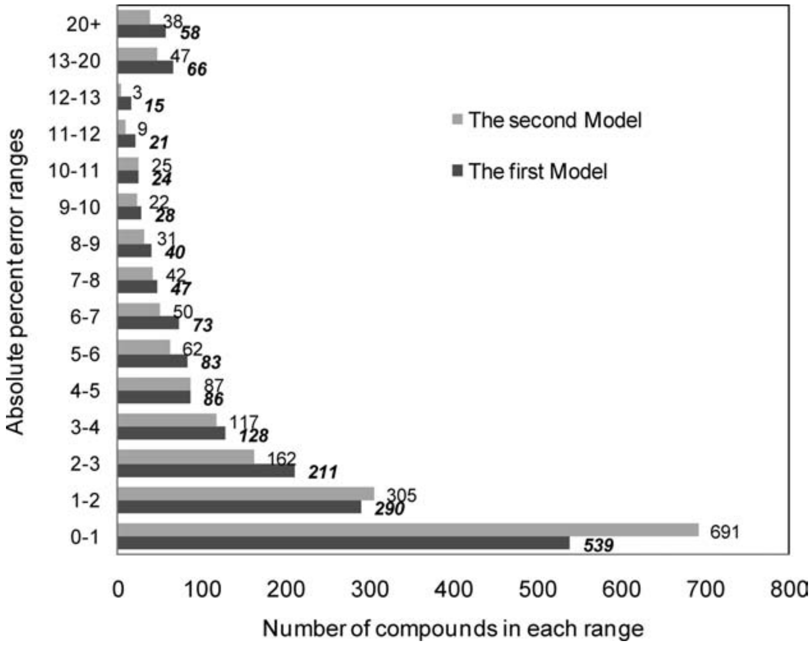


Figure 17. Deviation ranges of the results of the two developed models over all of the investigated compounds for acentric factors.

main reasons for existence of such outliers. First, the model is not able to represent or predict the properties of a portion of compounds with high accuracy. For instance, the developed model cannot estimate and calculate the small and simple molecules properties accurately because the majority of GC values are zero for the related compounds (e.g., methane). Therefore, the input parameters treated as the input parameters of the model cannot handle enough information to calculate and estimate the chemical structure of the related compound. The second reason is that we may suspect that corresponding experimental values are not accurate or may be somehow erroneous because of the existing difficulties in experimental measurements especially those where complex chemical structures are involved. For further investigation of the reliability of such data, we have pursued the following procedure:

1. Eliminating the outlier data points from the investigated experimental values.[12]
2. Developing a new ANN-GC model for calculation/estimation of the remaining properties values.
3. Prediction (estimation) of the eliminated outlier data points values using the new developed model for further checking the reliability of these values.

Table 3. Statistical Parameters of the Presented Models for Critical Pressure

statistical parameter	value	value
Training Set	First Model	Second Model
R^2	0.987	0.999
absolute average deviation	2.5 %	1.0 %
standard deviation error	20.1	19.2
mean square error	5.5	0.4
N	1373	1358
Validation Set		
R^2	0.978	0.998
absolute average deviation	2.3 %	1.3 %
standard deviation error	13.0	20.5
mean square error	3.8	0.8
N	171	169
Test Set		
R^2	0.986	0.999
absolute average deviation	2.1 %	1.0 %
standard deviation error	21.1	22.7
mean square error	6.5	0.5
N	171	169
Training + Validation + Test Set		
R^2	0.986	0.999
absolute average deviation	2.5 %	1.1 %
standard deviation error	19.6	19.7
mean square error	5.4	0.4
N	1715	1696

Table 4. Statistical Parameters of the Presented Models for Critical Volumes

statistical parameter	value	value
Training Set	First Model	Second Model
R^2	0.998	0.999
absolute average deviation	2.2 %	1.4 %
standard deviation error	0.3	0.3
mean square error	0	0
N	1375	1366
Validation Set		
R^2	0.998	0.998
absolute average deviation	1.8 %	1.4 %
standard deviation error	0.3	0.3
mean square error	0	0
N	171	170
Test Set		
R^2	0.998	0.999
absolute average deviation	2.3 %	1.2 %
standard deviation error	0.3	0.3
mean square error	0	0
N	171	170
Training + Validation + Test Set		
R^2	0.998	0.999
absolute average deviation	2.2 %	1.4 %
standard deviation error	0.3	0.3
mean square error	0	0
N	1717	1706

The results of the new developed model are shown in Figures 6 to 9. Figures 10 to 13 show the new predicted results for the eliminated outlier set from the main data set. More detailed results including the absolute deviations of the represented and predicted properties using the first and the second models are reported in Figures 14 to 17. Besides, the statistical parameters of both kinds of the models for each property are reported in Tables 2 to 5. As can be observed, the new model leads to the absolute deviation exceeding 20 % for none of the compounds from the new sets of data excluding the outliers for critical temperatures and pressures, 1 compound for critical volume, and 38 compounds for acentric factor. The average absolute deviations of this new adjusted model are about 1.1 % for P_c, 0.9 % for T_c, 1.4 % for V_c, and 3.7 % for the acentric factor while they are about 2.5 % for P_c, 1.5 % for T_c, 2.2 % for V_c, and 5.3 % for the acentric factor with model adjusted on all data (doubtful data included). All of the calculated and estimated results applying both of the models, the number of occurrences of the functional groups in all of investigated pure compounds, and the absolute deviations of the obtained results have been presented as Supporting Information. Tables 6 to 9 also illustrate the deviations of the final models over each chemical families.

Table 5. Statistical Parameters of the Presented Models for Acentric Factors

statistical parameter	value	value
Training Set	First Model	Second Model
R^2	0.987	0.992
absolute average deviation	5.5 %	3.7 %
standard deviation error	0.3	0.3
mean square error	0	0
N	1369	1353
Validation Set		
R^2	0.990	0.993
absolute average deviation	4.8 %	3.6 %
standard deviation error	0.3	0.3
mean square error	0	0
N	170	169
Test Set		
R^2	0.990	0.988
absolute average deviation	4.4 %	3.6 %
standard deviation error	0.3	0.3
mean square error	0	0
N	170	169
Training + Validation + Test Set		
R^2	0.987	0.992
absolute average deviation	5.3 %	3.7 %
standard deviation error	0.3	0.3
mean square error	0	0
N	1709	1691

It is inferred from the results that some of the predicted (estimated) physical property values for outlier (eliminated) data points bring about higher deviations (over 20 %) than the other deviation ranges even in the new developed model. We can refer to two possible reasons for this concept. At the first place, existence of the outlier points may lead the model not to be as accurate as possible. In other words, the obtained model considering the outlier points for "Training" process may not be predictive for the investigated chemical compounds in "Test (prediction)" process. Second, selecting the cut off value of the outliers (here it is 20 %) is so significant in developing the new (second) model. To the best of our knowledge, no definite procedure for selecting the cutoff value has been proposed in the

Table 6. Absolute Average Deviation Ranges of the Model Results for Critical Temperatures

family	AAD %
1-alkenes	1.2
2,3,4-alkenes	0.8
acetates	1.0
aldehydes	0.6
aliphatic ethers	0.6
alkylcyclohexanes	1.2
alkylcyclopentanes	0.7
alkynes	0.7
anhydrides	0.2
aromatic alcohols	1.0
aromatic amines	0.5
aromatic carboxylic acids	1.7
aromatic chlorides	0.5
aromatic esters	1.0
C, H, Br compounds	0.3
C, H, F compounds	0.5
C, H, I compounds	0.5
C, H multihalogen compounds	0.4
C, H, NO_2 compounds	0.9
C1/C2 aliphatic chlorides	1.3
C3 and higher aliphatic chlorides	0.5
cycloaliphatic alcohols	0.7
cycloalkanes	1.9
cycloalkenes	0.6
dialkenes	1.1
dicarboxylic acids	0.8
dimethylalkanes	1.4
diphenyl/polyaromatics	1.4
elements	2.5
epoxides	0.4
ethyl and higher alkenes	1.1
formates	0.5
inorganic acids	0.3
inorganic gases	1.2
inorganic halides	0.2
isocyanates/diisocyanates	0.4
ketones	0.7
mercaptans	1.0
methylalkanes	1.1
methylalkenes	0.8
multiring cycloalkanes	0.3
N-alcohols	0.7
N-aliphatic acids	0.5
N-aliphatic primary amines	0.3
N-alkanes	1.1
N-alkylbenzenes	0.5
naphthalenes	1.4
nitriles	1.3
nitroamines	1.6
organic salts	0.1
organic/inorganic compounds	3.4
other aliphatic acids	0.8
other aliphatic alcohols	1.2
other aliphatic amines	1.1
other alkanes	1.9
other alkylbenzenes	0.9
other amines, imines	0.4
other condensed rings	1.2
other ethers/diethers	1.0
other hydrocarbon rings	0.8
other inorganics	2.2
other monoaromatics	1.1
other polyfunctional C, H, O	2.1
other polyfunctional organics	0.0
other saturated aliphatic esters	1.3
peroxides	0.2
polyfunctional acids	1.3
polyfunctional amides/amines	0.5
polyfunctional C, H, N, halide, (O)	0.8
polyfunctional C, H, O, halide	0.6
polyfunctional C, H, O, N	0.2
polyfunctional C, H, O, S	0.3
polyfunctional esters	1.0
polyfunctional nitriles	1.1
polyols	1.4
propionates and butyrates	0.8
silanes/siloxanes	1.2
sulfides/thiophenes	0.4
terpenes	1.1
unsaturated aliphatic esters	1.0

Table 7. Absolute Average Deviation Ranges of the Model Results for Critical Pressures

family	AAD %
1-alkenes	1.1
2,3,4-alkenes	1.1
acetates	0.8
aldehydes	1.4
aliphatic ethers	1.5
alkylcyclohexanes	2.2
alkylcyclopentanes	1.8
alkynes	0.9
anhydrides	0.1
aromatic alcohols	1.8
aromatic amines	0.6
aromatic carboxylic acids	0.1
aromatic chlorides	2.2
aromatic esters	0.8
C, H, BR compounds	0.3
C, H, F compounds	0.3
C, H, I compounds	0.3
C, H multihalogen compounds	0.4
C, H, NO_2 compounds	1.3
C1/C2 aliphatic chlorides	0.6
C3 and higher aliphatic chlorides	0.5
cycloaliphatic alcohols	0.1
cycloalkanes	2.1
cycloalkenes	1.1
dialkenes	0.7
dicarboxylic acids	0.4
dimethylalkanes	2.4
diphenyl/polyaromatics	2.7
elements	0.3
epoxides	0.2
ethyl and higher alkenes	2.4
formates	0.9
inorganic acids	0.2
inorganic bases	0.5
inorganic gases	0.7
inorganic halides	1.5
isocyanates/diisocyanates	0.5
ketones	0.7
mercaptans	0.7
methylalkanes	1.1
methylalkenes	1.8
multiring cycloalkanes	0.7
N-alcohols	1.2
N-aliphatic acids	2.1
N-aliphatic primary amines	1.6
N-alkanes	1.3
N-alkylbenzenes	0.6
naphthalenes	2.6
nitriles	0.8
nitroamines	0.0

Table 7. Continued

family	AAD %
organic salts	0.0
organic/inorganic compounds	1.1
other aliphatic acids	1.6
other aliphatic alcohols	1.1
other aliphatic amines	1.7
other alkanes	3.5
other alkylbenzenes	2.0
other amines, imines	0.4
other condensed rings	1.4
other ethers/diethers	1.0
other hydrocarbon rings	0.9
other inorganics	0.7
other monoaromatics	1.0
other polyfunctional C, H, O	1.0
other polyfunctional ORGANICS	0.0
other saturated aliphatic esters	2.2
peroxides	0.2
polyfunctional acids	0.5
polyfunctional amides/amines	0.2
polyfunctional C, H, N, halide, (O)	0.2
polyfunctional C, H, O, halide	0.3
polyfunctional C, H, O, N	0.1
polyfunctional C, H, O, S	0.1
polyfunctional esters	0.9
polyfunctional nitriles	0.8
polyols	0.9
propionates and butyrates	1.4
silanes/siloxanes	2.8
sulfides/thiophenes	0.7
terpenes	0.8
unsaturated aliphatic esters	0.8

literature. In this work, we select this value based on the percent error ranges, after which the absolute deviations of the results from experimental values begin to soar gradually in comparison with other deviation ranges.

A significant point, which should be taken into account regarding the developed model, is that the uncertainties of the applied experimental values affect the model calculated/predicted results. However, these kinds of numerical models (QSPR and ANN-GC ones) are generally developed with the assumptions that the experimental data are considered error-free.[68] One should consider the uncertainties of the experimental data when using a develop model (especially the models, which are developed based on large data sets).

The mat file (MATLAB file format) of the new obtained ANN containing all the parameters of the model and the instruction for running the program have been presented as Supporting Information. The results obtained with the new obtained ANN-GC model prove it is an accurate and comprehensive method to calculate/estimate the critical properties and acentric factors of chemical compounds.

4. CONCLUSION

In this study, a group contribution-based model was presented for representation (calculation) and prediction (estimation) of critical properties and acentric factors of about 1700 chemical compounds. The models are the result of a combination of FFANNs and GC. The required parameters of the models are the numbers of occurrences of different functional groups in each investigated molecule. It should be noted that only few of these functional groups are simultaneously present in a particular molecule. Therefore, computation of the required parameters

Table 8. Absolute Average Deviation Ranges of the Model Results for Critical Volumes

family	AAD %
1-alkenes	1.4
2,3,4-alkenes	1.2
acetates	1.1
aldehydes	1.7
aliphatic ethers	1.0
alkylcyclohexanes	1.1
alkylcyclopentanes	2.0
alkynes	1.0
anhydrides	0.3
aromatic alcohols	2.8
aromatic amines	1.6
aromatic carboxylic acids	0.4
aromatic chlorides	1.4
aromatic esters	2.3
C, H, BR compounds	0.6
C, H, F compounds	1.1
C, H, I compounds	1.7
C, H multihalogen compounds	0.9
C, H, NO_2 compounds	1.8
C1/C2 aliphatic chlorides	0.6
C3 and higher aliphatic chlorides	0.5
cycloaliphatic alcohols	0.2
cycloalkanes	2.4
cycloalkenes	1.4
dialkenes	1.0
dicarboxylic acids	2.7
dimethylalkanes	1.7
diphenyl/polyaromatics	0.9
elements	5.5
epoxides	1.8
ethyl and higher alkenes	1.0
formates	1.9
inorganic acids	2.5
inorganic bases	0.8
inorganic gases	3.3
inorganic halides	0.5
isocyanates/diisocyanates	0.3
ketones	1.1
mercaptans	1.1
methylalkanes	2.0
methylalkenes	1.5
multiring cycloalkanes	0.2
N-alcohols	2.6
N-aliphatic acids	1.4
N-aliphatic primary amines	0.5
N-alkanes	0.9
N-alkylbenzenes	1.2
naphthalenes	2.7
nitriles	1.5
nitroamines	0.1
organic salts	0.4
organic/inorganic compounds	4.0
other aliphatic acids	1.7

Table 8. Continued

family	AAD %
other aliphatic alcohols	1.0
other aliphatic amines	2.1
other alkanes	2.3
other alkylbenzenes	1.3
other amines, imines	0.8
other condensed rings	2.0
other ethers/diethers	1.1
other hydrocarbon rings	1.3
other inorganics	4.8
other monoaromatics	1.4
other polyfunctional C, H, O	1.2
other polyfunctional organics	0.2
other saturated aliphatic esters	2.0
peroxides	0.6
polyfunctional acids	1.8
polyfunctional amides/amines	1.1
polyfunctional C, H, N, halide, (O)	1.3
polyfunctional C, H, O, halide	0.9
polyfunctional C, H, O, N	0.8
polyfunctional C, H, O, S	0.6
polyfunctional esters	1.0
polyfunctional nitriles	4.2
polyols	1.4
propionates and butyrates	1.1
silanes/siloxanes	1.7
sulfides/thiophenes	0.8
terpenes	1.1
unsaturated aliphatic esters	0.7

from chemical structure of any molecule is simple and straightforward. For developing the model, the experimental properties values from DIPPR[32] containing many of the pure compounds from various chemical families were applied. As a consequence, reliable and comprehensive models were developed to calculate/estimate the desired properties of many of pure compounds although there are still some limitations. Although these models have wide ranges of applicability, their prediction capabilities are restricted to the compounds, which are similar to those implied in its development. Consequently, the application of the obtained models for compounds that are chemically different to those investigated must be limited to rough estimations of their molecular diffusivity. However, the presented models may be conveniently used as a kind of technique to test the reliability of the experimental data reported in the literature for compounds with chemical structures similar to those already taken into account. It was found that experimental values of critical properties and acentric factors for some of the chemical compounds involved in former database are doubtful and must be considered with higher uncertainties than other experimental values.[32] After removing doubtful data, more satisfactory models have been obtained. In the final analysis, the obtained FFANN group contribution models can be considered as accurate ones for both representing the investigated properties and the predictions for compounds of similar chemical structure and also to point out unreliable data.

Table 9. Absolute Average Deviation Ranges of the Model Results for Acentric Factors

family	AAD %
1-alkenes	2.5
2,3,4-alkenes	3.2
acetates	3.2
aldehydes	3.3
aliphatic ethers	4.1
alkylcyclohexanes	9.8
alkylcyclopentanes	7.9
alkynes	3.4
anhydrides	0.3
aromatic alcohols	4.7
aromatic amines	1.5
aromatic carboxylic acids	0.5
aromatic chlorides	4.4
aromatic esters	2.6
C, H, BR compounds	0.9
C, H, F compounds	2.0
C, H, I compounds	4.6
C, H multihalogen compounds	2.3
C, H, NO_2 compounds	1.6
C1/C2 aliphatic chlorides	4.4
C3 and higher aliphatic chlorides	3.1
cycloaliphatic alcohols	0.7
cycloalkanes	14.9
cycloalkenes	5.8
dialkenes	11.1
dicarboxylic acids	1.1
dimethylalkanes	3.6
diphenyl/polyaromatics	3.7
elements	76.7
epoxides	1.3
ethyl and higher alkenes	3.3
formates	4.0
inorganic acids	8.7
inorganic bases	0.7
inorganic gases	11.4
inorganic halides	5.4
isocyanates/diisocyanates	4.6
ketones	3.2
mercaptans	2.2
methylalkanes	7.7
methylalkenes	3.5
multiring cycloalkanes	6.0
N-alcohols	2.8
N-aliphatic acids	3.5
N-aliphatic primary amines	2.2
N-alkanes	5.9
N-alkylbenzenes	2.4
naphthalenes	2.8
nitriles	3.0
nitroamines	1.3
organic salts	0.3
organic/inorganic compounds	17.6
other aliphatic acids	2.8

Table 9. Continued

family	AAD %
other aliphatic alcohols	4.3
other aliphatic amines	3.6
other alkanes	3.1
other alkylbenzenes	4.6
other amines, imines	2.6
other condensed rings	4.1
other ethers/diethers	2.1
other hydrocarbon rings	4.4
other inorganics	9.3
other monoaromatics	2.3
other polyfunctional C, H, O	3.8
other polyfunctional organics	0.2
other saturated aliphatic esters	3.9
peroxides	1.2
polyfunctional acids	0.8
polyfunctional amides/amines	1.2
polyfunctional C, H, N, halide, (O)	1.1
polyfunctional C, H, O, halide	1.5
polyfunctional C, H, O, N	0.5
polyfunctional C, H, O, S	0.5
polyfunctional esters	2.4
polyfunctional nitriles	1.2
polyols	2.7
propionates and butyrates	5.3
silanes/siloxanes	5.0
sulfides/thiophenes	1.7
terpenes	3.4
unsaturated aliphatic esters	2.2

APPENDIX

Instructions for Running the Program. The model is very easy to apply. Just drag and drop the mat file (freely available as Supporting Information) into the MATLAB environment (any version) workspace. One can follow the below example to get a response from the model step by step:

Assume that one is willing to predict the critical temperature of 2-methyl-1-pentene using the developed model. First of all, the group-contribution parameters should be defined from chemical structure of 2-methyl-1-pentene (refer to the Supporting Information). Later, drag and drop the mat file; the following commands should be entered in MATLAB workspace:

```
GC = [2  2 0 0 0 0 0 0 0 1 0 1 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 2 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0 0 ...
      0  0 0 0 0 0 0 0 0 0 0 0 0 0];
```

Tc_est=sim(net,GC′); 500.87 K; where its experimental value is equal to 505 (ARD % = 0.8 %).

ASSOCIATED CONTENT

Ⓢ Supporting Information. 16 files (12 excel spreadsheets and a zip file containing four mat files) including the schematic list of the functional groups, number of occurrences of all of the functional groups in the investigated compounds, the calculated and predicted values of critical properties and acentric factors, the obtained results, the chemical formulas of all of the investigated compounds, and the absolute relative deviations of the two presented model results with/without considering of outliers (as xls files) and the computer (mat) program developed based on the proposed model (as a mat file). 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.

Funding Sources

A.E. wishes to thank MINES ParisTech for providing him a Ph.D. scholarship.

REFERENCES

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

(2) 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.

(3) Godavarthy, S. S.; Robinson, R. L., Jr.; Gasem, K. A. M. Improved structure—property relationship models for prediction of critical properties. *Fluid Phase Equilib.* **2008**, *264*, 122–136.

(4) Ambrose, D.; Young, C. L. Vapor-liquid critical properties of elements and compounds 1. An introduction survey. *J. Chem. Eng. Data* **1995**, *40*, 345–357.

(5) Sola, D.; Ferri, A.; Banchero, M.; Manna, L.; Sicardi, S. QSPR prediction of N-boiling point and critical properties of organic compounds and comparison with a group-contribution method. *Fluid Phase Equilib.* **2008**, *263*, 33–42.

(6) Kudchadker, A. P.; Zwolinski, B. J. Vapor pressures and boiling points of normal alkanes C21 to C100. *J. Chem. Eng. Data* **1966**, *11*, 253–255.

(7) Lydersen, A. L. *Estimation of Critical Properties of Organic Compounds*, Eng. Exp. Stn. Rep. 3; University of Wisconsin College Engineering: Madison, WI, 1955.

(8) Joback, K. G.; Reid, R. C. Estimation of pure component properties from group contributions. *Chem. Eng. Commun.* **1987**, *57*, 233–243.

(9) Marrero, J.; Gani, R. Group-contribution based estimation of pure component properties. *Fluid Phase Equilib.* **2001**, *183*−*184*, 183–208.

(10) Kudchadker, A. P.; Holcomb, W. D.; Zwollinski, B. J. Correlation studies on vapor pressures and critical properties for isomeric alkanes. *J. Chem. Eng. Data* **1968**, *13*, 182–188.

(11) Thodos, G. Critical constants of saturated aliphatic hydrocarbons. *AIChE J.* **1955**, *1*, 168–173.

(12) Thodos, G. Critical constants of unsaturated aliphatic hydrocarbons. *AIChE J.* **1955**, *1*, 165–168.

(13) Thodos, G. Critical constants of naphthenic hydrocarbons. *AIChE J.* **1956**, *2*, 508–513.

(14) Thodos, G. Critical constants of aromatic hydrocarbons. *AIChE J.* **1957**, *3*, 428–431.

(15) Ambrose, D.; Ghiasse, N. B. Vapor pressures and critical temperatures and critical pressures of some alkanoic acids: C1 to C10. *J. Chem. Thermodyn.* **1987**, *19*, 505–519.

(16) Wilson, G. M.; Jasperson, L. V. Critical constants T_c and P_c, estimation based on zero, first and second order methods, Paper given at AIChE Spring National Meeting, New Orleans, LA, USA, February 25−29, 1996.

(17) Morejón-Marrero, J.; Pardillo-Fontdevila, E. Estimation of pure compound properties using group-interaction contributions. *AIChE J.* **1999**, *45*, 615–621.

(18) Lydersen, A. L.; Tsochev, V. Critical temperatures and pressures and high temperature vapour pressures of seven alcohols. *Chem. Eng. Technol.* **1990**, *13*, 125–130.

(19) Dalmazzone, D.; Salmon, A.; Guella, S. A second order group contribution method for the prediction of critical temperatures and enthalpies of vaporization of organic compounds. *Fluid Phase Equilib.* **2006**, *242*, 29–42.

(20) Joback, K. G.; A unified approach to physical property estimation using multivariate statistical techniques. M.Sc. Thesis, MIT, Cambridge, MA, 1984.

(21) Constantinou, L.; Gani, R. New group-contribution method for the estimation of properties of pure compounds. *AIChE J.* **1994**, *10*, 1697–1710.

(22) Somayajulu, G. R. Estimation procedures for critical constants. *J. Chem. Eng. Data* **1989**, *34*, 106–120.

(23) Kreglewski, A. P.; Zwolinski, B. J. A new relation for physical properties of n-alkanes and n-alkyl compounds. *J. Phys. Chem.* **1961**, *65*, 1050–1052.

(24) Gasem, K. A. M. Binary vapor−liquid equilibrium for carbon dioxide + heavy normal paraffins. Ph.D. Dissertation, Oklahoma State University, Stillwater, OK, 1986.

(25) Gasem, K. A. M.; Ross, C. H.; Robinson, R. L., Jr. Prediction of ethane an CO_2 solubilities in heavy paraffins using generalized-parameter Soave and Peng−Robinson equations of state. *Can. J. Chem. Eng.* **1993**, *71*, 805–816.

(26) Tsonopoulos, C.; Tan, Z. The critical constants of normal alkanes from methane to polyethylene. II. Application of the Flory theory. *Fluid Phase Equilib.* **1993**, *83*, 127–138.

(27) Magoulas, K.; Tassios, D. Thermophysical properties of n-alkanes from C1 to C20 and their prediction for higher ones. *Fluid Phase Equilib.* **1990**, *56*, 119–140.

(28) Teja, A. S.; Smith, R. L. Critical properties of thermally unstable substances from mixture data. *AIChE J.* **1987**, *33*, 1560–1562.

(29) Marano, J. J.; Holder, G. D. General equation for correlating the thermophysical properties of n-paraffins, n-olefins, and other homologous series. 2. Asymptotic behavior correlations for PVT properties. *Ind. Eng. Chem. Res.* **1997**, *36*, 1895–1907.

(30) Morgan, D. L.; Kobayashi, R. Direct vapor pressure measurements of ten n-alkanes in the C 10 -C 28 range. *Fluid Phase Equilib.* **1994**, *97*, 211.

(31) Kontogeorgis, G. M.; Tassios, D. P. Critical constants and acentric factors for long-chain alkanes suitable for corresponding states applications: A critical review. *Chem. Eng. J.* **1997**, *66*, 35–49.

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

(33) 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.* **2010**, *49*, 10149–10152.

(34) Hou, T. J.; Wang, J. M. Structure - ADME relationship: still a long way to go. *Exp. Opin. Drug Metab. Toxicol.* **2008**, *4*, 759–771.

(35) 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.

(36) Gharagheizi, F.; Sattari, M.; Tirandazi, B. Prediction of crystal lattice energy using enthalpy of sublimation: A group contribution-based model. *Ind. Eng. Chem. Res* **2011**, *50*, 2482–2486.

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

(38) 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.

(39) 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.

(40) 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.

(41) 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.

(42) 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.

(43) 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.

(44) 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.

(45) 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.

(46) 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.

(47) 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.

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

(49) 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.

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

(51) Gharagheizi, F.; Tirandazi, B.; Barzin, R. Estimation of aniline point temperature of pure hydrocarbons: A quantitative structure-property relationship approach. *Ind. Eng. Chem. Res.* **2009**, *48*, 1678–1682.

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

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

(56) Gharagheizi, F. A new neural network quantitative structure-property relationship for prediction of θ (Lower Critical Solution Temperature) of polymer solutions. *e-Polym.* **2007**, No. 114.

(57) 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.

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

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

(60) 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.

(61) Gharagheizi, F.; Abbasi, R. A new neural network group contribution method for estimation of upper flash point of pure chemicals. *Ind. Eng. Chem. Res.* **2010**, *49*, 12685–12695.

(62) 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.

(63) 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.

(64) Gharagheizi, F.; Eslamimanesh, A.; Mohammadi, A. H.; Richon, D. Use of artificial neural network-group contribution method to determine surface tension of pure compounds. *J. Chem. Eng. Data* **2011**, doi: 10.1021/je2001045.

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

(66) Hagan, M.; Demuth, H. B.; Beale, M. H. *Neural Network Design*; International Thomson Publishing: Tampa, FL, 2002.

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

(68) Kazakov, A.; Muzny, C. D.; Diky, V.; Chirico, R. D.; Frenkel, M. Predictive correlations based on large experimental datasets: Critical constants for pure compounds. *Fluid Phase Equilib.* **2010**, *298*, 131–142.