An accurate model for prediction of autoignition temperature of pure compounds

Farhad Gharagheizi *

Saman Energy Giti Co., Postal Code 3331619636, Tehran, Iran

ARTICLE INFO

ABSTRACT

Accurate prediction of pure compounds autoignition temperature (AIT) is of great importance. In this study, the Artificial Neural Network-Group Contribution (ANN-GC) method is applied to evaluate the AIT of pure compounds. 1025 pure compounds from various chemical families are investigated to propose a comprehensive and predictive model. The obtained results show the squared correlation coefficient of 0.984, root mean square error of 15.44 K, and average percent error of 1.6% for the experimental values.

1. Introduction

If the temperature of a flammable gas–air mixture is uniformly raised, it eventually reaches to a value, at which combustion occurs. For the range of flammable mixtures, there is a mixture composition which has the lowest ignition temperature. This minimum temperature is called autoignition temperature (AIT) or spontaneous ignition temperature (SIT). The AIT is defined as the lowest temperature, at which a substance will produce hot-flame ignition in air at atmospheric pressure without the aid of an external energy source such as a spark or flame [1].

At the AIT, the rate of heat evolved by exothermic oxidation reaction overbalances the rate at which heat is lost to the surroundings and causes ignition. The AIT is dependent not only on the chemical and physical properties of the substance but also on the method and instrument employed for its determination such as the volume and the material of the used vessel, test pressure, and oxygen concentration [1].

Due to importance of the AIT of fuels, time-consuming, and laborious processes of experimental determination of the AIT, theoretical computation of the AIT is of great interest. There have been several models so far presented for this purpose. For instance, Suzuki [2] developed some equations to estimate the AIT using a number of several molecular-based parameters and physical properties. The model showed average absolute deviation of 4.5% and squared correlation coefficient of 0.9. The correlation was devel-

oped using a dataset for 250 chemical compounds. Furthermore, the correlation was unable to estimate the AIT of 23 compounds with enough accuracy. In the case that these compounds are added to the calculations, the squared correlation coefficient decreases to 0.85 and the average absolute deviation increased to 5.4%. Another attempt has been done by Tetteh et al. [3], who tried to modify the Suzuki's correlation by presenting an Artificial Neural Network instead of the correlation. Therefore, they used the parameters previously applied by Suzuki as the inputs of their models. They reported the error of 30 °C using the same correlation used by Suzuki. Mitchel and Jurs [4] developed quantitative structure–property relationship (QSPR) for estimation of the AIT. They applied a dataset including 327 heterogeneous organic compounds to develop their model. Moreover, they stated that their attempt to model all the compounds together was unsuccessful. Hence, these researchers presented several models each of them appropriate for calculation of the AIT of a portion of the compounds in the dataset. Another QSPR approach was developed by Kim et al. [5] using a dataset including 200 compounds. Their model showed that the square of the correlation coefficient (R^2) for the AIT of the 157-member training set was 0.920, and the root mean square error was 25.876. The squared correlation coefficient and root mean square error of their model for a 43-member prediction set were 0.910 and 28.968, respectively. Albahri and George [6] and Albahri [7] presented two very similar models for estimation of AIT. It seems the first one was a bit better. The model showed squared correlation coefficient of 0.98 and average error of 2.8% over a data set including 490 compounds.

Recently, some other models have been presented for estimation of AIT property. It seems that these models are not more general and accurate than the model presented by Albahri and

* Current address: Saman Energy Giti Co., Postal Code 3331619636 Tehran, Iran.
Fax: +98 21 66957784.
E-mail addresses: fghara@ut.ac.ir, fghara@gmail.com

George [6]. For instance, we can refer to the two models proposed by Pan et al. [8,9,10]. For the first one, they used a small data set including 192 compounds to develop a QSPR-based support vector machine type method. Although the square correlation coefficient of the model obtained was 0.984, the considerable difference between the leave-one-out parameter and squared correlation parameter showed its poor predictive power (Q_{LOO}). Later, they presented an Artificial Neural Network-Group Contribution model using a dataset including 116 hydrocarbons. The average absolute error and root mean square error of the model were 21.6 °C and 31.1, respectively. Finally, using a data set containing 446 organic compounds, Pan et al. [10] developed a QSPR-based support vector machine type mode to estimate the AIT. The squared correlation coefficient and root mean square of the model were 0.9 and 36.86 °C. A recent model presented by Chen et al. [11]. They used the classic group contribution method to calculate the AIT. The squared correlation coefficient of the model was 0.85.

One of the main problems that have been mentioned in majority of the presented models is that the AIT values are hard to be correlated with chemical structure of compounds. This issue can be easily inferred from the low accuracy of the previously presented models. Another issue is related to the number of compounds that have been used in developing the previous models. Even though, there are experimental data for much more than 500 compounds available in the literature, but none of the aforementioned researchers applied such wide ranges of data for developing their models.

The main aim of this study is to present a new comprehensive model for calculation/estimation of AIT using a large dataset containing more than 1000 compounds.

2. Procedure

2.1. Data set preparation

DIPPR 801 [12] database has been found especial applications in developing new methods for prediction of physical properties because it contains a large number of pure compounds as well as their physical properties. In order to provide a data set for calculation of the AIT, 1025 pure compounds from 78 different chemical groups were investigated and their related AITs were considered for the study. These compounds are presented as supplementary material.

2.2. Development of a new group contribution

Having defined the dataset, the chemical structures of all 1025 compounds were analyzed. Consequently, 146 functional groups were found to be more efficient for calculation/estimation of the AIT of pure compounds. The functional groups used in this study are presented in Table 1 . Besides, their numbers of occurrences in pure compounds used in this work are also extensively presented as supplementary materials. These chemical groups are used as the proposed method parameters.

The next calculation step and perhaps the most significant one is to search for a relationship between the chemical functional groups and the AIT of chemical compounds. The simplest method for this purpose is assumption of existence of a multi-linear relationship between these groups and the desired property (here is the AIT of pure compounds) [13–14]. This technique is a similar method used in the most of classical group contribution methods. 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 investigated.

Artificial Neural Networks are extensively used in various scientific and engineering problems [13–41] e.g. estimations of physical and chemical properties of different pure compounds [28–41]. These capable mathematical tools are generally applied to study the complicated systems [13–41]. The theoretical explanations about Neural Networks can be found elsewhere [42]. Using the Neural Network toolbox of the MATLAB software (Mathworks® Inc.), a three layer Feed Forward Artificial Neural Network (FFANN) is developed for the problem. The capabilities of this kind of ANNs have been demonstrated in the previous works [13–15,28–41].

All the 146 functional groups and also the AIT 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 AITs for output parameters. Later, the main dataset is divided into three new sub-datasets including the "training" set, the "validation (optimization)" set, and the "test (prediction)" 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 main dataset into three sub-datasets is performed randomly. For this purpose, about 80%, 10%, and 10% of the main dataset are randomly selected for the "training" set (821 compounds), the "validation" set (102 compounds), and the "test" set (102 compounds). The effect of the allocation percent of the three sub-datasets from the data of main dataset on the accuracy of the ANN model has been studied elsewhere [43]. The typical structure of a three layer FFANN is schematically presented in Fig. 1.

This type of ANN contains three layers; the input layer, the hidden layer, and the output layer. It consists of n neurons. The parameter n is one of the main parameters of the FFANN and should be obtained by post-optimization after generating the base-structure of the FFNN. The mathematic formulation of this strategy can be written as follows:

$$\text{output}(i) = W_2 \times \tanh(W_1 \times \text{input}(i) + b_1) + b_2 \qquad (1)$$

where i refers to the data point, the parameter W_1 is the weight relating the first layer to the second layer, W_2 is the weight relating the second to the third or output layer, b_1 denotes the bias to the second layer, and b_2 stands for the bias of third or output layer. W_1, W_2, b_1, and b_2 should be determined by a process called "training".

Generally, to generate a FFNN, there is a need to know some parameters of the proposed problem. The required parameters are the number of inputs (nip), number of outputs (nop), and the number of objects ($nobj$) i.e. number of data points with known input parameter (ip) and corresponding output parameter (op). For instance, in the problem in this study with 146 input parameters (functional groups) and one output parameter (the AIT) with 821 data points (the training set compounds), the nip, nop, and $nobj$ are 146, 1, and 821, respectively.

As a matter of fact, generating an ANN model is determination of the weight matrices and bias vectors [42]. As shown in Fig. 1, there are two weight matrices and two bias vectors in a three layer FFANN: W_1 and W_2, b_1 and b_2 [13–42]. 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 (predicted AITs) and the target values (experimental AIT values). This minimization is performed by Levenberg–Marquardt (LM) [42] 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

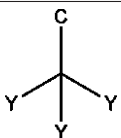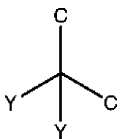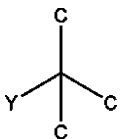**Table 1**
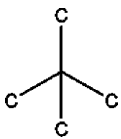Functional groups used to develop the model.
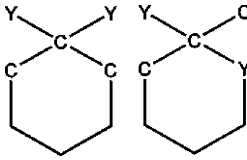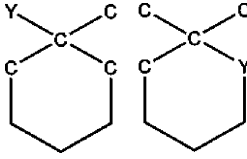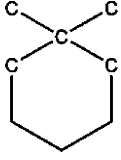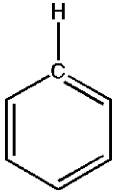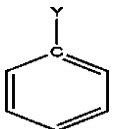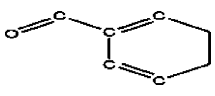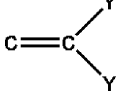
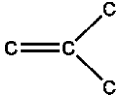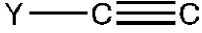| No.		Functional groups	Comments
1	AIT001		Terminal primary C(sp3) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH$_2$, etc.)
2	AIT002		Total secondary C(sp3) Y = H or any heteroatom
3	AIT003		Total tertiary C(sp3) Y = H or any heteroatom
4	AIT004		Total quaternary C(sp3)
5	AIT005		Ring secondary C(sp3) Y = H or any heteroatom
6	AIT006		Ring tertiary C(sp3) Y = H or any heteroatom
7	AIT007		Ring quaternary C(sp3)
8	AIT008	Sum of all the carbons belonging to any aromatic and heteroaromatic structure	Aromatic C(sp2)
9	AIT009		Unsubstituted benzene C(sp2)
10	AIT010		Substituted benzene C(sp2) Y = carbon or any heteroatom
11	AIT011		Non-aromatic conjugated C(sp2)
12	AIT012		Terminal primary C(sp2) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH$_2$, etc.)

Table 1 (*Continued*)
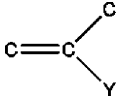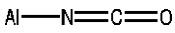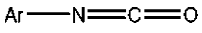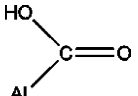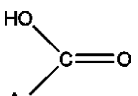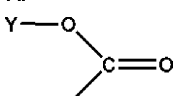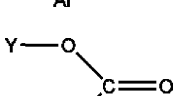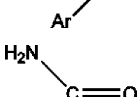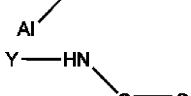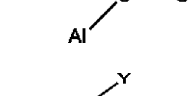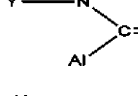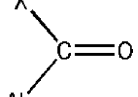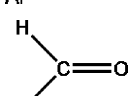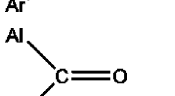
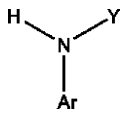| No.		Functional groups	Comments
13	AIT013		Aliphatic secondary C(sp2) Y = H or any heteroatom
14	AIT014		Aliphatic tertiary C(sp2)
15	AIT015	Y—C≡C	Terminal C(sp) Y = any terminal atom or heteroaromatic group (i.e. H, X, OH, NH$_2$, etc.)
16	AIT016	Y—C≡C	Non-terminal C(sp) Y = C or any non-terminal heteroatom
17	AIT017	Al—N=C=O	Isocyanates (aliphatic)
18	AIT018	Ar—N=C=O	Isocyanates (aromatic)
19	AIT019		Carboxylic acids (aliphatic)
20	AIT020		Carboxylic acids (aromatic)
21	AIT021		Esters (aliphatic) Y = Ar or Al (not H) Al = H or aliphatic group linked through C
22	AIT022		Esters (aromatic) Y = Al or Ar
23	AIT023		Primary amides (aliphatic) Al = H or aliphatic group linked through C
24	AIT024		Secondary amides (aliphatic) Y = Ar or Al (not H, not C=O) Al = H or aliphatic group linked through C
25	AIT025		Tertiary amides (aliphatic) Y = Ar or Al (not H, not C=O) Al = H or aliphatic group linked through C
26	AIT026		Acyl halogenides (aliphatic)
27	AIT027		Aldehydes (aromatic)
28	AIT028		Ketones (aliphatic)
29	AIT029		Carbonate (-thio) derivatives (Y = O or S)

Table 1 (*Continued*)
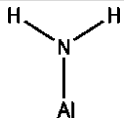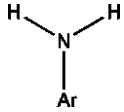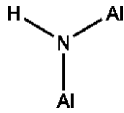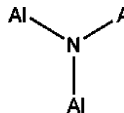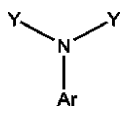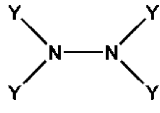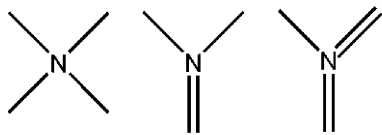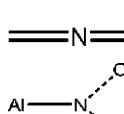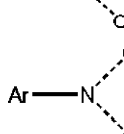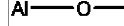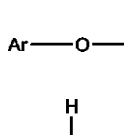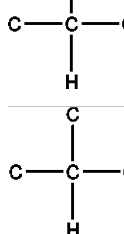
No.		Functional groups	Comments
30	AIT030		Primary amines (aliphatic) Al = aliphatic group linked through C (not C=O)
31	AIT031		Primary amines (aromatic)
32	AIT032		Secondary amines (aliphatic) Al = aliphatic group linked through C (not C=O)
33	AIT033		Secondary amines (aromatic) Y = Ar or Al (not C=O)
34	AIT034		Tertiary amines (aliphatic) Al = aliphatic group linked through C (not C=O)
35	AIT035		Tertiary amines (aromatic) Y = Ar or Al (not C=O)
36	AIT036		N Hydrazines Y = C or H
37	AIT037	N≡C—Al	Nitriles (aliphatic)
38	AIT038		Positive charged N
39	AIT039		Nitro groups (aliphatic) Al = H or aliphatic group linked through carbon
40	AIT040		Nitro groups (aromatic) Al = aromatic group linked through carbon
41	AIT041	Al—O—H	Hydroxyl groups Al = aliphatic group linked through any atom
42	AIT042	Ar—O—H	Aromatic hydroxyls Ar = aromatic group linked through any atom
43	AIT043		Primary alcohols
44	AIT044		Secondary alcohols

Table 1 (*Continued*)

No.		Functional groups	Comments
45	AIT045		Tertiary alcohols
46	AIT046		Ethers (aliphatic) Al = aliphatic group linked through C (not C=O, not C#N)
47	AIT047		Ethers (aromatic) Y = Ar or Al (not C=O, not C#N)
48	AIT048		Anhydrides (thio-) Y = O or S
49	AIT049		Thiols
50	AIT050		Sulfides
51	AIT051		Disulfides
52	AIT052		Sulfates (thio-/dithio-) (Y = O or S)
53	AIT053		Phosphates/thiophosphates (Y = O or S)
54	AIT054		CH_2RX
55	AIT055		CHR_2X
56	AIT056		R=CHX
57	AIT057		R=CRX
58	AIT058		$CHRX_2$
59	AIT059		CR_2X_2
60	AIT060		CRX_3

Table 1 (*Continued*)

No.		Functional groups	Comments
61	AIT061	Ar—X	X on aromatic ring
62	AIT062		X on exo-conjugated C
63	AIT063		Aziridines
64	AIT064		Oxiranes
65	AIT065		Thiranes
66	AIT066		Furanes
67	AIT067		Thiophenes
68	AIT068		Pyridines
69	AIT069	Sum of the hydrogens linked to all of the Os and Ns in the molecul	Donor atoms for H-bonds (N and O)
70	AIT070	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)
71	AIT071	CH_3R/CH_4	
72	AIT072	CH_2R_2	
73	AIT073	CHR_3	
74	AIT074	CR_4	
75	AIT075	CH_3X	
76	AIT076	CH_2RX	
77	AIT077	CH_2X_2	
78	AIT078	CHR_2X	
79	AIT079	$CHRX_2$	
80	AIT080	CHX_3	
81	AIT081	CR_3X	
82	AIT082	CRX_3	
83	AIT083	$=CH_2$	
84	AIT084	$=CHR$	
85	AIT085	$=CR_2$	
86	AIT086	$=CHX$	
87	AIT087	$=CRX$	
88	AIT088	$=CX_2$	
89	AIT089	#CH	
90	AIT090	#CR/R=C=R	
91	AIT091	R–CH–R	
92	AIT092	R–CR–R	
93	AIT093	R–CX–R	
94	AIT094	R–CH–X	
95	AIT095	R–CR–X	
96	AIT096	R–CH..X	
97	AIT097	R–CR..X	
98	AIT098	Al–CH=X	
99	AIT099	Ar–CH=X	
100	AIT100	Al–C(=X)–Al	
101	AIT101	R–C(=X)–X/R–C#X/X=C=X	
102	AIT102	X–C(=X)–X	
103	AIT103	H^a attached to $C^0(sp3)$ no X attached to next C	
104	AIT104	H^a attached to $C^1(sp3)/C^0(sp2)$	

Table 1 (*Continued*)

No.		Functional groups	Comments
105	AIT105	H[a] attached to C^2(sp3)/C^1(sp2)/C^0(sp)	
106	AIT106	H[a] attached to C^3(sp3)/C^2(sp2)/C^3(sp2)/C^3(sp)	
107	AIT107	H attached to heteroatom	
108	AIT108	H attached to alpha-C[b]	
109	AIT109	H[a] attached to C^0(sp3) with 1X attached to next C	
110	AIT110	H[a] attached to C^0(sp3) with 2X attached to next C	
111	AIT111	H[a] attached to C^0(sp3) with 3X attached to next C	
112	AIT112	alcohol	
113	AIT113	phenol/enol/carboxyl OH	
114	AIT114	=O	
115	AIT115	Al–O–Al	
116	AIT116	Al–O–Ar/Ar–O–Ar/R..O..R/R–O–C=X	
117	AIT117	O–[c]	
118	AIT118	R–O–O–R	
119	AIT119	Al–NH_2	
120	AIT120	Al_2–NH	
121	AIT121	Al_3–N	
122	AIT122	Ar–NH_2/X–NH_2	
123	AIT123	Ar–NH–Al	
124	AIT124	Ar–NAl_2	
125	AIT125	RCO–N</>N–X=X	
126	AIT126	Ar_2NH/Ar_3 N/Ar_2N–Al/R..N..R[c]	
127	AIT127	R#N/R=N–	
128	AIT128	R–N–R[d]/R–N–X	
129	AIT129	Ar–NO_2/R–N(–R)–O[e]/RO–NO	
130	AIT130	Al–NO_2	
131	AIT131	F[a] attached to C^3(sp3)	
132	AIT132	F[a] attached to C^2(sp2)-C^4(sp2)/C^1(sp)/C^4(sp3)/X	
133	AIT133	Cl[a] attached to C^1(sp3)	
134	AIT134	Cl[a] attached to C^2(sp3)	
135	AIT135	Cl[a] attached to C^3(sp3)	
136	AIT136	Cl[a] attached to C^1(sp2)	
137	AIT137	Cl[a] attached to C^2(sp2)-C^4(sp2)/C^1(sp)/C^4(sp3)/X	
138	AIT138	Br[a] attached to C^1(sp3)	
139	AIT139	Br[a] attached to C^1(sp2)	
140	AIT140	I[a] attached to C^1(sp3)	
141	AIT141	R–SH	
142	AIT142	R_2S/RS–SR	
143	AIT143	R=S	
144	AIT144	R–SO_2–R	
145	AIT145	>Si<	
146	AIT146	X_3–P=X (phosphate)	

Explanations: 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. The formal oxidation number of a carbon atom equals the sum of the conventional bond orders with electronegative atoms; the C–N bond order in pyridine may be considered as 2 while we have one such bond and 1.5 when we have two such bonds; the C..X bond order in pyrrole or furan may be considered as 1.

[b] An alpha-C may be defined as a C attached through a single bond with –C=X, –C#X, –C–X.

[c] Pyrrole-type structure.

[d] Pyridine-type structure.

[e] Pyridine N-oxide type structure.

[41–43] is most-widely used algorithm for training due to being robust and accurate enough to deal with the considered system [13–41].

In most cases, the number of neurons in the hidden layer (n) is fixed. Therefore, the main goal is to produce a ANN model, which is able to predict the target values as accurately as expected. This step is repeated till 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 [13–43].

3. Results and discussion

An optimized Feed Forward Artificial Neural Network has been obtained using the aforementioned procedure for prediction of the

AIT of 1025 pure compounds. For this 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 $n = 10$. 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 [28,42]. Therefore, the developed three-layer FFANN has the structure of 146-10-1.

The *mat* file (MATLAB® file format) of the obtained Artificial Neural Network containing all the parameters of the model and the instruction for running the program are freely available upon request to the authors. The predicted AITs are shown in Fig. 2 in comparison with the experimental values [12].

The statistical parameters of the model are presented in Table 2. These results show that the squared correlation coefficient, average
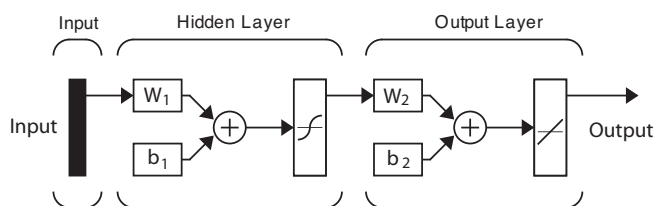
Fig. 1. The schematic structure of the three-layer Feed Forward Artificial Neural Network used in this study.
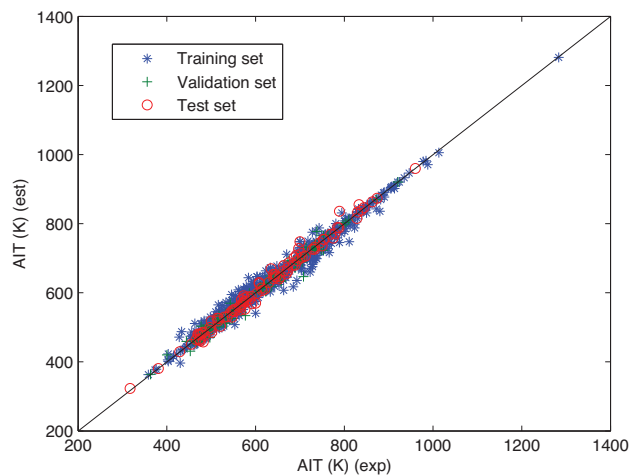
Fig. 2. Comparison between the calculated/estimated results of the model and experimental values[12] of the AIT of the investigated pure chemical compounds.

percent error, and root mean square error of the model over the training set, the validation set, the test set and the main data set are 0.984, 0.980, 0.986, 0.984, 1.6%, 1.6%, 1.6%, 1.6%, 15.4, 16.25, 14.99, 15.44 K, respectively.

The average percent errors of the model for the investigated chemical families of pure compounds are shown in Table 3. It was found from the results that the model can accurately predict the AIT of different classes of pure compounds. Based on the results, the maximum percent error is related to the diethyl ether which is

Table 2
Statistical parameters of the presented models.

Statistical parameter	Value
Training set	
R^2	0.984
Average percent error	1.60%
Standard deviation	119.11
Root mean square error	15.4
n	821
Validation set	
R^2	0.980
Average percent error	1.60%
Standard deviation	110.8
Root mean square error	16.25
n	102
Test set	
R^2	0.986
Average percent error	1.60%
Standard deviation	125.38
Root mean square error	14.99
n	102
Training + validation + test set	
R^2	0.984
Average percent error	1.60%
Standard deviation	118.95
Root mean square error	15.44
n	1025

Table 3
The average absolute percent errors of the model for all of the investigated chemical families.

| ID	Family	AAD%
1	1-Alkenes	2.90
2	2,3,4-Alkenes	3.39
3	Acetates	1.75
4	Aldehydes	1.61
5	Aliphatic ethers	2.77
6	Alkylcyclohexanes	1.81
7	Alkylcyclopentanes	4.15
8	Alkynes	0.35
9	Anhydrides	0.43
10	Aromatic alcohols	1.41
11	Aromatic amines	0.39
12	Aromatic carboxylic acids	0.22
13	Aromatic chlorides	0.17
14	Aromatic esters	1.54
15	C, H, BR compounds	0.34
16	C, H, F compounds	0.19
17	C, H, I compounds	0.36
18	C, H, multihalogen compounds	0.07
19	C, H, NO_2 compounds	2.38
20	C1/C2 aliphatic chlorides	0.33
21	C3 & higher aliphatic chlorides	0.95
22	Cycloaliphatic alcohols	0.47
23	Cycloalkanes	5.38
24	Cycloalkenes	2.47
25	Dialkenes	1.24
26	Dicarboxylic acids	0.62
27	Dimethylalkanes	2.31
28	Diphenyl/polyaromatics	1.84
29	Epoxides	0.78
30	Ethyl & higher alkenes	2.99
31	Formates	3.68
32	Inorganic bases	0.38
33	Inorganic gases	0.63
34	Isocyanates/diisocyanates	0.23
35	Ketones	2.11
36	Mercaptans	0.27
37	Methylalkanes	2.21
38	Methylalkenes	2.62
39	Multiring cycloalkanes	0.97
40	N-Alcohols	1.26
41	N-Aliphatic acids	1.58
42	N-Aliphatic primary amines	1.41
43	N-Alkanes	2.13
44	N-Alkylbenzenes	2.82
45	Naphthalenes	2.24
46	Nitriles	1.03
47	Nitroamines	3.04
48	Organic salts	0.24
49	Organic/inorganic compounds	5.00
50	Other aliphatic acids	1.34
51	Other aliphatic alcohols	1.98
52	Other aliphatic amines	1.20
53	Other alkanes	2.48
54	Other alkylbenzenes	3.32
55	Other amines, imines	0.77
56	Other condensed rings	0.55
57	Other ethers/diethers	1.49
58	Other hydrocarbon rings	0.82
59	Other inorganics	0.06
60	Other monoaromatics	2.26
61	Other polyfunctional C, H, O	2.10
62	Other polyfunctional organics	0.10
63	Other saturated aliphatic esters	1.16
64	Peroxides	0.30
65	Polyfunctional acids	0.38
66	Polyfunctional amides/amines	0.52
67	Polyfunctional C, H, N, halide, (O)	0.22
68	Polyfunctional C, H, O, halide	0.57
69	Polyfunctional C, H, O, N	0.31
70	Polyfunctional C, H, O, S	0.47
71	Polyfunctional esters	1.58
72	Polyfunctional nitriles	0.11
73	Polyols	1.99
74	Propionates and butyrates	1.93
75	Silanes/siloxanes	1.58
76	Sulfides/thiophenes	0.55
77	Terpenes	1.87
78	Unsaturated aliphatic esters	0.80
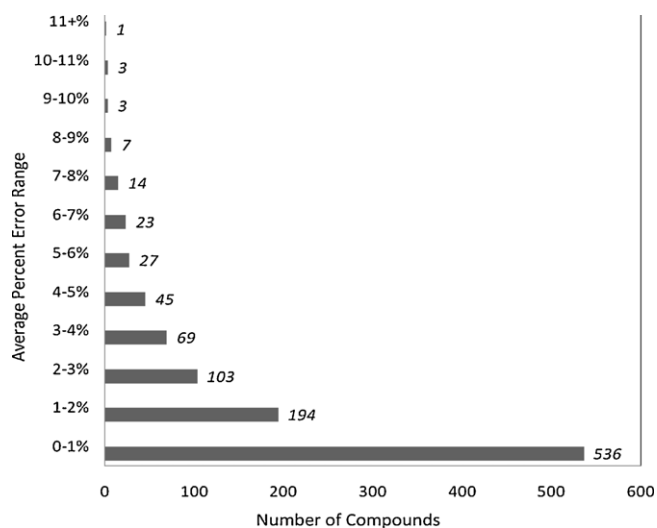
Fig. 3. The absolute percent error ranges of the model over all of the investigated compounds.

12.5%. According to the DIPPR 801, the uncertainties of the experimental AIT values reported for the diethyl ether have not been reported in the original source, so it is probable that this value is not as accurate as other compounds reported values. Additionally, it is the only compound that the corresponding value of AIT is predicted with more than 11% deviation from experimental value. The quality of the model is better understood using a chart as in Fig. 3 that shows the average percent error ranges. As can be easily seen, more than half of the main dataset, which is larger than the most comprehensive dataset that have ever been used in previous studies [6], is predicted with lower than 1% deviation from experimental value. This accomplishment confirms the capabilities of the model.

4. Conclusion

A group contribution-based model was presented for prediction of the autoignition temperature (AIT) of pure compounds. The model is the result of a combination of group contributions and Feed Forward Artificial Neural Networks. The parameters of the model are the numbers of occurrences of 146 functional groups in each investigated molecule. It should be noted that majority of these 146 functional groups are not simultaneously available in a particular molecule, so computation of these functional parameters from chemical structure of any molecule is simple. For developing the model, a large data set involving 1025 pure compounds were used. Consequently, this model can be used to predict the AIT of every regular compound but with some limitations. 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 molecular diffusivity of these kinds of compounds.

Finally, the average absolute deviation of the model results from experimental values[12] demonstrates the accuracy of the presented model.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jhazmat.2011.02.014.

References

[1] ASTM International, ASTM Standard Test Method E659-78, The American Society for Testing and Materials, West Conshohocken, PA, 2000.
[2] T. Suzuki, Quantitative structure–property relationships for auto-ignition temperatures of organic compounds, Fire Mater. 18 (1994) 81–88.
[3] J. Tetteh, E. Metcalfe, S.L. Howells, Optimisation of radial basis and back-propagation neural networks for modelling auto-ignition temperature by quantitative-structure property relationships, Chemom. Intell. Lab. Syst. 32 (1996) 177–191.
[4] B.E. Mitchel, P.C. Jurs, Prediction of autoignition temperatures of organic compounds from molecular structure, J. Chem. Inf. Comput. Sci. 37 (1997) 538–547.
[5] Y.K. Kim, S.K. Le, J.H. Kim, J.S. Kim, K.T. No, Prediction of autoignition temperatures (AITs) for hydrocarbons and compounds containing heteroatoms by the quantitative structure–property relationship, J. Chem. Soc. Perk Trans. 2 (2002) 2087–2092.
[6] T.A. Albahri, R.S. George, Artificial neural network investigation of the structural group contribution method for predicting pure components auto ignition temperature, Ind. Eng. Chem. Res. 42 (2003) 5708–5714.
[7] T.A. Albahri, Flammability characteristics of pure hydrocarbons, Chem. Eng. Sci. 58 (2003) 3629–3641.
[8] Y. Pan, J. Jiang, R. Wang, H. Cao, Advantages of support vector machine in QSPR studies for predicting auto-ignition temperatures of organic compounds, Chemom. Intell. Lab. Syst. 92 (2008) 169–178.
[9] Y. Pan, J. Jiang, R. Wang, H. Cao, J. Zhao, Prediction of auto-ignition temperatures of hydrocarbons by neural network based on atom-type electrotopological-state indices, J. Hazard. Mater. 157 (2008) 510–517.
[10] Y. Pan, J. Jiang, R. Wang, H. Cao, R. Cui, Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine, J. Hazard. Mater. 164 (2009) 1242–1249.
[11] C.C. Chen, H.J. Liaw, Y.Y. Kuo, Prediction of autoignition temperatures of organic compounds by the structural group contribution approach, J. Hazard. Mater. 162 (2009) 746–762.
[12] Project 801, Evaluated Process Design Data, Public Release Documentation, Design Institute for Physical Properties (DIPPR), American Institute of Chemical Engineers (AIChE), 2006.
[13] F. Gharagheizi, R. Abbasi, B. Tirandazi, Prediction of Henry's law constant of organic compounds in water from a new group-contribution-based model, Ind. Eng. Chem. Res. 49 (2010) 10149–10152.
[14] F. Gharagheizi, R. Abbasi, A New Neural Network Group Contribution Method for Estimation of Upper Flash Point of Pure Chemicals, Ind. Eng. Chem. Res. 49 (2010) 12658–12695.
[15] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, Artificial neural network modeling of solubilities of 21 commonly used industrial solid compounds in supercritical carbon dioxide, Ind. Eng. Chem. Res. 50 (2011) 221–226.
[16] A. Chouai, S. Laugier, D. Richon, Modeling of thermodynamic properties using neural networks: application to refrigerants, Fluid Phase Equilib. 199 (2002) 53–62.
[17] L. Piazza, G. Scalabrin, P. Marchi, D. Richon, Enhancement of the extended corresponding states techniques for thermodynamic modelling. I. Pure fluids, Int. J. Refrig. 29 (2006) 1182–1194.
[18] G. Scalabrin, P. Marchi, L. Bettio, D. Richon, Enhancement of the extended corresponding states techniques for thermodynamic modelling. II. Mixtures, Int. J. Refrig. 29 (2006) 1195–1207.
[19] A. Chapoy, A.H. Mohammadi, D. Richon, Predicting the hydrate stability zones of natural gases using artificial neural networks, Oil Gas Sci. Technol.: Rev. IFP 62 (2007) 701–706.
[20] A.H. Mohammadi, D. Richon, Hydrate phase equilibria for hydrogen + water and hydrogen + tetrahydrofuran + water systems: predictions of dissociation conditions using an artificial neural network algorithm, Chem. Eng. Sci. 65 (2010) 3352–3355.
[21] A.H. Mohammadi, D. Richon, Estimating sulfur content of hydrogen sulfide at elevated temperatures and pressures using an artificial neural network algorithm, Ind. Eng. Chem. Res. 47 (2008) 8499–8504.
[22] A.H. Mohammadi, D. Richon, A mathematical model based on artificial neural network technique for estimating liquid water–hydrate equilibrium of water–hydrocarbon system, Ind. Eng. Chem. Res. 47 (2008) 4966–4970.
[23] A.H. Mohammadi, W. Afzal, D. Richon, Determination of critical properties and acentric factors of petroleum fractions using artificial neural networks, Ind. Eng. Chem. Res. 47 (2008) 3225–3232.
[24] A.H. Mohammadi, D. Richon, Use of artificial neural networks for estimating water content of natural gases, Ind. Eng. Chem. Res. 46 (2007) 1431–1438.
[25] A.H. Mohammadi, J.F. Martínez-López, D. Richon, Determining phase diagrams of tetrahydrofuran + methane, carbon dioxide or nitrogen clathrate hydrates using an artificial neural network algorithm, Chem. Eng. Sci. 65 (2010) 6059–6063.
[26] M. Mehrpooya, A.H. Mohammadi, D. Richon, Extension of an artificial neural network algorithm for estimating sulfur content of sour gases at elevated temperatures and pressures, Ind. Eng. Chem. Res. 49 (2010) 439–442.

[27] A.H. Mohammadi, V. Belandria, D. Richon, 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. 65 (2010) 4302–4305.

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

[29] F. Gharagheizi, New neural network group contribution model for estimation of lower flammability limit temperature of pure compounds, Ind. Eng. Chem. Res. 48 (2009) 7406–7416.

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

[31] F. Gharagheizi, B. Tirandazi, R. Barzin, Estimation of aniline point temperature of pure hydrocarbons: a quantitative structure–property relationship approach, Ind. Eng. Chem. Res. 48 (2009) 1678–1682.

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

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

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

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

[36] F. Gharagheizi, A new neural network quantitative structure–property relationship for prediction of θ (Lower Critical Solution Temperature) of polymer solutions, e-Polymers (2007), Article no 114.

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

[38] F. Gharagheizi, A chemical structure-based model for estimation of upper flammability limit of pure compounds, Energy Fuel 27 (2010) 3867–3871.

[39] A. Vatani, M. Mehrpooya, F. Gharagheizi, Prediction of standard enthalpy of formation by a QSPR model, Int. J. Mol. Sci. 8 (2007) 407–432.

[40] M. Mehrpooya, F. Gharagheizi, A Molecular approach for prediction of sulfur compounds solubility parameters, phosphorus sulfur and silicon and related elements, Phosphorus Sulfur 185 (2010) 204–210.

[41] S.A. Kalogirou, Artificial neural networks in renewable energy systems applications: a review, Renew. Sust. Energy Rev. 5 (2001) 373–401.

[42] M. Hagan, H.B. Demuth, M.H. Beale, Neural Network Design, International Thomson publishing, 2002.

[43] F. Gharagheizi, QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN, Comput. Mater. Sci. 40 (2007) 159–165.