Korean J. Chem. Eng., 34(2), 425-439 (2017)
DOI: 10.1007/s11814-016-0271-7

pISSN: 0256-1115
eISSN: 1975-7220

Implementation of soft computing approaches for prediction of physicochemical properties of ionic liquid mixtures

Saeid Atashrouz*, Hamed Mirshekar, Abdolhossein Hemmati-Sarapardeh***, Mostafa Keshavarz Moraveji****,†, and Bahram Nasernejad******

*Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Mahshahr Campus, Mahshahr, Iran
**Iran Polymer and Petrochemical Institute (IPPI), Tehran, Iran
***Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
****Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez 424, P. O. Box 15875-4413, Tehran, Iran
(*Received 4 May 2016 • accepted 30 September 2016*)

Abstract−The main objective of this study was to develop soft computing approaches for prediction of physicochemical properties of IL mixtures including: density, heat capacity, thermal conductivity, and surface tension. The proposed models in this study are based on support vector machine (SVM), least square support vector machines (LSSVM), and group method of data handling type polynomial neural network (GMDH-PNN) systems. To find the LSSVM and SVM adjustable parameters, genetic algorithm (GA) as a meta-heuristic algorithm was utilized. The results showed that LSSVM is more robust and reliable for prediction of physicochemical properties of IL mixtures. The proposed GA-LSSVM model provides average absolute relative deviations of 0.38%, 0.18%, 0.77% and 1.18% for density, heat capacity, thermal conductivity, and surface tension, respectively, which demonstrates high accuracy of the model for prediction of physicochemical properties of IL mixtures.

Keywords: Physicochemical Properties, Ionic Liquid, GMDH-PNN, LSSVM, SVM

INTRODUCTION

Ionic liquids (ILs) are a new class of liquids which maintain their liquid state for wide ranges of temperature [1]. ILs have attracted the attention in green chemistry due to their interesting and unique physicochemical properties, such as very low vapor pressure, ease of recycle, high solvating capacity for polar and nonpolar compounds, nonvolatility, and high thermal, chemical, and electrochemical stability [2]. In addition, physicochemical properties of ILs are tunable [3]. Because, various combinations of anions and cations can lead to creation of new ILs and consequently with tunable physicochemical properties [4]. ILs have been considered as novel solvents in several fields and applications, including chemical synthesis and catalytic reactions, lithium ion batteries biotechnology, nanotechnology, fuel cells, dye-sensitized solar cells, extraction, and separation technology [5,6]. To employ ILs and their mixtures for industrial applications, knowledge of their physicochemical properties is important [7]. In recent years many experimental studies have focused on measurement of physicochemical properties of mixtures of IL, including density, surface tension, heat capacity, thermal and electrical conductivity, viscosity, refractive index, and speed of sound [8-10]. However, it is time consuming and costly to measure the physicochemical properties of mixtures of IL for wide ranges of temperature and mole fractions. Consequently, development of predictive models plays a key role in employing these mixtures for various applications.

Some studies have proposed models for correlating physicochemical properties of IL mixtures. Alvarez et al. proposed a COSMO-SAC model for correlating density data of three IL mixtures [11]. Deviation for their model is lower than 3%. Hosseini et al. employed a perturbed hard-sphere equation of state (EOS) for prediction of density data for 14 IL mixtures with mean absolute relative deviation (MARD) of 0.38% [12]. Yousefi developed extended Tao and Mason EOS for estimation of density data for 13 systems with MARD of 1.69% [13]. Andreatta et al. used the Redlich-Kister polynomial equation for correlating the surface tension of six systems with standard deviation lower than 0.1 [14]. Lin et al. used the Redlich-Kister equation for correlation of heat capacity of two systems with MARD=0.1% [15]. Atashrouz et al. proposed a modified two suffix Gibbs free energy model and coupled it with Eyring's relation to develop a model for estimation of viscosity of IL mixtures [6]. They considered a wide range of experimental data (5512 data points) form 122 systems, and an MARD of 2.07% was obtained for their model. Note that for other physicochemical properties of IL mixtures such as speed of sound, refractive index, and electrical conductivity, the Redlich-Kister equation has been used [16,17]. A literature review showed that although physicochemical properties of IL mixtures have been investigated extensively, few predictive models have been developed for the aim of prediction of these properties. Despite good accuracy of the aforementioned

†To whom correspondence should be addressed.
E-mail: moraveji@aut.ac.ir

models, they have not the ability to predict a physicochemical property when interaction parameters of the system are not available. Therefore, these models can be regarded only as "correlative" models which have limited applications. In this regard, development of "predictive" models with reasonable accuracy has a great importance for IL mixtures.

Over the last decade, artificial intelligence systems such as artificial neural network (ANN) have been widely used for modeling various problems [18-21]. ANNs are predictive tools with high flexibility to learn the behavior of highly nonlinear problems [22]. ANN has been used for prediction of physicochemical properties of IL mixtures by Huang et al., and the developed model exhibited good performance [7]. Although ANNs are powerful tools with high degree of flexibility, they are involved with some disadvantages. ANN has a complex mathematical structure which is not able to provide an explicit analytical equation for relating the output parameter to input parameters [23]. In addition, it has a high number of adjustable parameters called weights and biases. Also, another disadvantage of ANNs is the overfitting issue which may occur when the ANN has a good performance for training data, but has poor ability to predict test data. In this regard, the use of alternative systems seems to be necessary.

Support vector machine (SVM) has attracted attention as a novel intelligent method which can be used for pattern recognition, classification and regression problems. SVM has a strong theoretical basis according to the statistical learning theory. In the SVM method, instead of using the traditional empirical risk minimization principle, which is commonly used for neural networks, structural risk minimization principle is implemented. This can significantly prevent over-fitting and under-fitting problems [24]. Furthermore, SVM is usually solved based on quadratic programming (QP) [25]. Because QP is a convex function, there is only one global and unique solution for a problem which has more benefits than neural networks with many local solutions [24]. More advantages of SVM in comparison to ANNs are provided in the literature [26-28] Hosseinzadeh and Hemmati-Sarapardeh developed an LSSVM model for prediction of viscosity of ternary mixtures containing ILs [29]. The MARD% for their model is 1.39, which demonstrates high capability of LSSVM for precise predicting viscosity of IL mixtures.

Another alternative for ANN can be the group method of data handling type polynomial neural network (GMDH-PNN) systems [30,31]. The main idea of GMDH is using feed-forward networks based on polynomial transfer functions. Such a network includes layers and nodes which can provide an explicit analytical model demonstrating relationship between input and output variables [2,32]. In our previous studies, we proposed GMDH-PNNs for modeling different systems [2,23,33,34]. GMDH-PNN model was implemented for estimation of the viscosity of nanofluids, and the model exhibited good accuracy (MARD=2.14%) [23]. Also, a GMDH-PNN model was developed for prediction of water activity in binary systems of water and poly(ethylene glycol) (PEG). The root mean square error for the GMDH-NN model was obtained 0.0348 [33]. Additionally, GMDH-PNN was also used for the prediction of surface tension of pure ILs, and the obtained MARD was 4.59% [34]. This method has been successfully applied for modeling thermal conductivity of pure ILs with an MARD of 1.79% [2]. Good performance of GMDH-PNN in the abovementioned studies, and especially in the studies about pure ILs, demonstrates that this system can be used as a potential candidate in the field of ILs. Consequently, it would be interesting to apply this method for modeling physicochemical properties of mixtures of IL.

In this study, SVM, least square SVM (LSSVM), and GMDH-PNN models as soft computing approaches are applied to predict physicochemical properties of IL mixtures, including density, heat capacity, thermal conductivity, and surface tension. The parameters of LSSVM and SVM models are optimized using genetic algorithm (GA). To this end, a wide range of IL mixtures were collected from variety of literature sources and the data points were randomly split into two sets as training (60% of the data) set and test set (40% of the data). After the models were developed, several statistical and graphical error analyses were used to evaluate the performance of the models in both of the training and testing sets.

METHODOLOGY

1. Data Acquisition

Physicochemical properties of mixtures are considered to be a function of temperature, mole fraction of IL and second compound, and physicochemical properties of pure components (density, surface tension, heat capacity and thermal conductivity). For instance, for prediction of density of a mixture, the input parameters are temperature of the system, mole fraction of IL and the second compound, density of pure IL and density of the second compound. Table 1 demonstrates characteristics of experimental data for density. 33 binary mixtures containing 1104 density data points were considered for model development. A total of 27 individual compounds were implemented, in which the ILs have various cations (imidazolium, pyridinium, ammonium, phosphonium) and anions (BF_4, BTI, C_6SO_4, C_8SO_4, $MeSO_4$, Gly, Br, NO_3, DCA). Characteristics of experimental data for heat capacity are tabulated in Table 2, in which nine binary mixtures containing 603 heat capacity data points have been considered. Totally, nine individual compounds have been used, in which the cations of ILs are imidazolium and pyridinium, and also anions include BF_4, $MeSO_4$, PF_6 and TfO. For the case of thermal conductivity, characteristics of experimental data are reported in Table 3, in which three binary mixtures containing 24 thermal conductivity data points have been considered. Totally five individual compounds have been used in these mixtures. Cations of ILs are imidazolium type, while anions include DMP, $EtSO_4$ and TfO. Table 4 demonstrates characteristics of experimental data for surface tension. 28 binary mixtures containing 573 surface tension data points have been used for model development. A total of 25 individual compounds have been implemented, in which the ILs have various cations (imidazolium, pyridinium, ammonium, phosphonium) and anions (BF_4, BTI, C_4SO_4, C_6SO_4, C_8SO_4, $MeSO_4$, $MeSO_3$, Gly, Br, NO_3, DCA, $EtSO_4$). The accuracy of the selected data for developing model was compared with those reported in other papers (for available cases). It was found the data are consistent and valid.

2. Statistical Assessments

With the aim of evaluating performance of the proposed models in this study, several statistical parameters including MARD%,

Table 1. List of the selected binary systems for modeling density

No.	Binary system	Temperature range/K	Density range/(g·cm^{-3})	Data points	Ref.
1	[C4MIm][PF6] (1)+[C4MIm][BF4] (2)	303.15	1.2082-1.3554	12	[44]
2	[C6MIm][BF4] (1)+[C2MIm][BF4] (2)	298.15	1.1058-1.2616	13	[44]
3	[C6MIm][Cl] (1)+[C6MIm][PF6] (2)	303.15-333.15	1.0776-1.2467	16	[45]
4	[C8MIm][Cl] (1)+[C8MIm][BF4] (2)	313.15-333.15	1.0098-1.0767	8	[45]
5	2-Butanol (1)+[MOA][BTI] (2)	298.15-313.15	0.8826-1.1008	30	[46]
6	2-Butanone (1)+[C4MIm][PF6] (2)	293.15-303.15	0.7949-1.3716	39	[47]
7	2-Butanone (1)+[Mmim][MeSO4] (2)	293.15-303.15	0.7949-1.3309	27	[47]
8	2-Propanol (1)+[MOA][BTI] (2)	298.15-313.15	0.8651-1.1005	30	[46]
9	Acetic acid (1)+[BMIM][SCN] (2)	293.15-313.15	1.0271-1.0740	60	[48]
10	Acetic acid (1)+[EMIM][EtSO4] (2)	293.15-313.15	1.0271-1.2403	60	[49]
11	Acetone (1)+[C4MIm][PF6] (2)	298.15	0.7848-1.3678	15	[50]
12	Acetonitrile (1)+[BMIM][SCN] (2)	293.15-313.15	0.7602-1.0727	60	[51]
13	Acetonitrile (1)+[EMIM][EtSO4] (2)	293.15-313.15	0.7602-1.2403	60	[51]
14	Dimethyl carbonate (1)+[C6MIm][PF6] (2)	298.15	1.0633-1.2937	13	[52]
15	Ethanol (1)+[C4MIm][BF4] (2)	298.15	0.7854-1.2012	14	[53]
16	Ethanol (1)+[Mmim][MeSO4] (2)	298.15	0.7854-1.3291	13	[54]
17	Ethyl acetate (1)+[C4MIm][PF6] (2)	293.15-303.15	0.8882-1.3716	39	[47]
18	Ethyl acetate (1)+[Mmim][MeSO4] (2)	293.15-303.15	0.8882-1.3309	21	[47]
19	Methanol (1)+[C4MIm][SCN] (2)	298.15-328.15	0.7578-1.0697	51	[55]
20	Methyl acetate (1)+[C4MIm][BF4] (2)	298.15	0.9159-1.2110	15	[56]
21	Methyl formate (1)+[C4MIm][BF4] (2)	298.15	0.9647-1.2110	15	[56]
22	n-Hexane (1)+[C8MIm][PF6] (2)	293.15-303.15	0.6504-1.2396	18	[57]
23	Propionic acid (1)+[BMIM][SCN] (2)	293.15-313.15	0.9717-1.0727	60	[48]
24	Propionic acid (1)+[EMIM][EtSO4] (2)	293.15-313.15	0.9717-1.2403	60	[49]
25	Water (1)+[C2MIm][BTI] (2)	298.19	1.5013-1.5193	8	[55]
26	Water (1)+[C4MPyr][BTI] (2)	298.21-323.21	1.3653-1.3953	15	[55]
27	Water (1)+[C6MIm][BF4] (2)	298.15	0.9980-1.1463	18	[57]
28	Water (1)+[C6MIm][Cl] (2)	298.15	0.9972-1.0423	11	[58]
29	Water (1)+[C8MIm][Cl] (2)	298.15-343.15	0.9757-1.0127	44	[58]
30	Water (1)+[empy][EtSO4] (2)	298.15-328.15	0.9857-1.2222	39	[4]
31	Water (1)+[N1114][BTI] (2)	293.15-343.15	1.3438-1.3995	88	[55]
32	Water (1)+[pDMIM][BF4] (2)	298.15-323.15	0.9881-1.2251	66	[59]
33	Water (1)+[C4(3-m)Py][BF4] (2)	293.15-318.15	0.9902-1.1886	66	[60]

Table 2. List of the selected binary systems for modeling heat capacity

No.	Binary system	Temperature range/K	Heat capacity range/(J·mol^{-1}K^{-1})	Data points	Ref.
1	Acetonitrile (1)+[C6MIm][BF4] (2)	283.15-323.15	90.03-445.8	90	[61]
2	Acetonitrile (1)+[C8MIm][BF4] (2)	283.15-323.15	90.03-513.3	99	[61]
3	Ethanol (1)+[bmpyr][BF4] (2)	293.15-318.15	111.3-399.3	96	[60]
4	Methanol (1)+[C6MIm][BF4] (2)	283.15-323.15	79.5-444	71	[62]
5	Methanol (1)+[C8MIm][BF4] (2)	283.15-323.15	78.6-514.2	72	[62]
6	Water (1)+[C4MIm][BF4] (2)	303.2-353.2	137.5-335.7	44	[63]
7	Water (1)+[C4MIm][MeSO4] (2)	303.2-353.2	140-355	43	[64,65]
8	Water (1)+[C4MIm][PF6] (2)	303.2-353.2	342.2-422.6	44	[63]
9	Water (1)+[C4MIm][TfO] (2)	303.2-353.2	149-398	44	[64]

Table 3. List of the selected binary systems for modeling thermal conductivity

No.	Binary system	Temperature range/K	Thermal conductivity range/(W·m^{-1}K^{-1})	Data points	Ref.
1	Methanol (1)+[Mmim][DMP] (2)	298.15	0.192-0.238	8	[66]
2	Water (1)+[C2mim][EtSO4] (2)	293	0.181-0.607	8	[42]
3	Water (1)+[C4MIm][TfO] (2)	293	0.147-0.607	8	[42]

Table 4. List of the selected binary systems for modeling surface tension

No.	Binary system	Temperature range/K	Surface tension range/(m·N·m^{-1})	Data points	Ref.
1	1-Butanol (1)+[C4MIm][BTI] (2)	298.15	23.80-32.80	13	[67]
2	1-Propanol (1)+[C4MIm][BTI] (2)	298.15	23.34-32.80	11	[67]
3	Acetonitrile (1)+[C2MIm][BTI] (2)	293.15-313.15	26.43-36.14	46	[68]
4	Acetonitrile (1)+[C4MIm][BTI] (2)	293.15-313.15	26.43-32.78	45	[69]
5	Dimethyl sulfoxide (1)+[C2MIm][BTI] (2)	293.15-313.15	35.10-43.36	45	[69]
6	Dimethyl sulfoxide (1)+[C4MIm][BTI] (2)	293.15-313.15	31.88-43.36	50	[69]
7	Ethanol (1)+[C2MIm][C6SO4] (2)	298.15	22.98-34.77	13	[70]
8	Ethanol (1)+[C2MIm][C8SO4] (2)	298.15	22.89-30.91	18	[70]
9	Ethanol (1)+[C4MIm][BF4] (2)	298.15	22.83-45.33	12	[53]
10	Ethanol (1)+[C6MIm][BF4] (2)	298.15	23.11-37.33	10	[53]
11	Ethanol (1)+[C8MIm][BF4] (2)	298.15	23.33-32.82	10	[53]
12	Methanol (1)+[C1MIm][MeSO4] (2)	298.15	22.51-59.80	9	[71]
13	Methanol (1)+[C2MIm][MeSO4] (2)	298.15	21.80-37.21	7	[72]
14	Tetrahydrofuran (1)+[C2MIm][BTI] (2)	293.15-308.15	25.38-36.14	44	[68]
15	Tetrahydrofuran (1)+[C4MIm][BTI] (2)	293.15-308.15	25.38-32.72	39	[69]
16	Water (1)+[C1MIm][MeSO4] (2)	296.8-298.1	64.70-71.10	10	[73]
17	Water (1)+[C2MIm][BF4] (2)	298.15	53.04-54.43	8	[53]
18	Water (1)+[C2MIm][C4SO4] (2)	298.15	39.08-45.93	27	[70]
19	Water (1)+[C2MIm][EtSO4] (2)	298.15	47.00-50.11	12	[74]
20	Water (1)+[C2MIm][MeSO3] (2)	300-303.3	57.30-71.60	27	[73]
21	Water (1)+[C4MIm][BF4] (2)	298.15	45.33-46.32	15	[53]
22	Water (1)+[C4MIm][Gly] (2)	283.15-328.15	41.50-46.60	60	[75]
23	Water (1)+[C4Py][NO3] (2)	298.15	37.50-51.50	15	[76]
24	Water (1)+[C6MIm][BF4] (2)	298.15	37.33-37.93	8	[53]
25	Water (1)+[N112(hoe)][Br] (2)	298.15	63.98-66.97	3	[71]
26	Water (1)+[N311(hoe)][Br] (2)	298.15	47.72-51.99	3	[71]
27	Water (1)+[P666(14)][BTI] (2)	298.1-342.3	26.88-30.05	6	[77]
28	Water (1)+[P666(14)][Dca] (2)	298.2-342.8	29.50-32.56	6	[77]

mean deviation (MD%), root mean square deviation (RMSD), relative deviation (RD%) and regression coefficient (R^2) are implemented as shown below:

$$\mathrm{MARD}(\%)=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{P^{calc,\,i}-P^{exp,\,i}}{P^{exp,\,i}}\right| \quad (1)$$

$$\mathrm{RMSD}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(P^{calc,\,i}-P^{exp,\,i})^2} \quad (2)$$

$$\mathrm{MD}(\%)=\frac{100}{N}\sum_{i=1}^{N}(P^{calc,\,i}-P^{exp,\,i}) \quad (3)$$

$$\mathrm{RD}(\%)=100\times\frac{P^{calc,\,i}-P^{exp,\,i}}{P^{exp,\,i}} \quad (4)$$

$$R^2=1-\left[\frac{\sum_{i=1}^{N}(P^{exp,\,i}-P^{calc,\,i})^2}{\sum_{i=1}^{N}(P^{exp,\,i}-\overline{P})^2}\right] \quad (5)$$

where N is the number of data points, and P is the output of the model (a physicochemical property) and also $\overline{P}$ is the average of experimental data points.

3. Support Vector Machine

The support vector machine can be employed for both regression and classification problems [25]. Based on the SVM method, any function f(x) can be written as follows:

$$f(x)=w^T\varphi(x)+b \quad (6)$$

where x is the input data which has a dimension of N×n, where N and n demonstrate the number of data points and input variables, respectively. Furthermore, w_T, $\varphi(x)$ and b are the transposed output layer, the kernel function and bias vector, respectively. Fig. 1

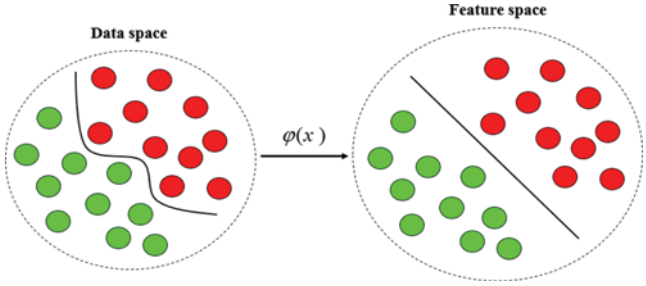

Fig. 1. The main idea of the SVM method.

demonstrates the main idea of SVM method for solution of problems. The primary objects on the right hand of Fig. 1 were rearranged based on Kernel function, which is called mapping (transformation). Then the mapped objects at the right hand of Fig. 1 are linearly separable. Consequently, instead of using a complex curve, a line can separate red and green objects, which leads to simplicity for solving problems.

To calculate w and b, the following cost function (CF) was developed by Vapnik [35]:

$$\mathrm{CF}=\frac{1}{2}w^{T}+c\sum_{k=1}^{N}(\xi_k-\xi_k^*) \tag{7}$$

where ξ and ξ^* stand for flexible variables which govern the training error and should be used to determine the allowed margin of error. Furthermore, c is considered as the adjustable parameter of the SVM, which is a factor that measures the trade-off between model's capacity to predict unseen data and training accuracy [24].

The following constraints are used for the above relation:

$$\begin{cases} y_k-w^{T}\varphi(x_k)-b\le\varepsilon+\xi_k, & k=1, 2, \ldots, N \\ w^{T}\varphi(x_k)+b-y_k\le\varepsilon+\xi_k^*, & k=1, 2, \ldots, N \\ \xi_k,\ \xi_k^*\ge 0 & k=1, 2, \ldots, N \end{cases} \tag{8}$$

where y_k and x_k are the kth output data and the kth input data, respectively. ε is a parameter representing the tube width of Vapnik's ε-insensitive loss function with the aim of controlling the regression precision and improving the model performance [24], which can be observed in Fig. 2.

To minimize the mentioned cost function, Lagrangian multipliers and Karush-Kuhn-Tucker (KKT) conditions are employed in Eq. (7), which leads to the following Lagrangian form of the problem [35]:

$$\begin{aligned} L(a, a^*)=&-\frac{1}{2}\sum_{k, l=1}^{N}(a_k-a_k^*)(a_l-a_l^*)K(x_k-x_l) \\ &-\varepsilon\sum_{k=1}^{N}(a_k-a_k^*)+\sum_{k=1}^{N}y_k(a_k-a_k^*) \end{aligned} \tag{9}$$

$$\sum_{k=1}^{N}(a_k-a_k^*)=0,\quad a_k, a_k^*\in[0, c] \tag{10}$$

$$K(x_k-x_l)=\varphi(x_k)^{T}\varphi(x_l)\quad k=1, 2, \ldots, N \tag{11}$$

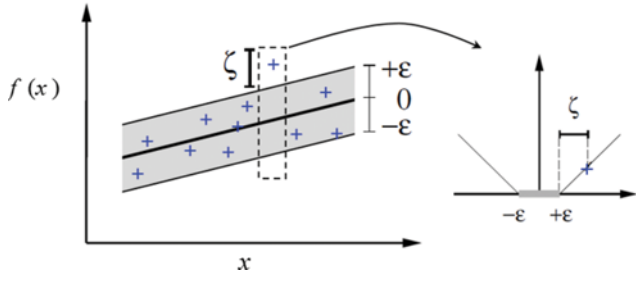

Fig. 2. (Left) Tube of ε-accuracy and points which cannot meet this accuracy, (Right) Vapnik ε-insensitive loss function for function estimation.

where a_k and a_k^* are the Lagrangian multipliers. Based on the above equations, the final mathematical form of SVM takes the following form [35]:

$$f(x)=\sum_{k, l=1}^{N}(a_k-a_k^*)K(x-x_k)+b \tag{12}$$

In the SVM method, quadratic programing (QP) is used for solving equations [24]. To reduce the complexity of calculations, recently the least square formulation of SVM (LSSVM) has been proposed with the idea of converting quadratic program into linear equality sets [36,37]. LSSVM is a simpler version of SVM but with good abilities of SVM. It is easier to implement and has shorter computing time. According to the LSSVM method, a linear set of equations should be solved for obtaining solutions of the model. The cost function (CF) of LSSVM has the following form [36]:

$$\mathrm{CF}=\frac{1}{2}w^{T}w+\frac{1}{2}\gamma\sum_{k=1}^{N}e_k^2 \tag{13}$$

This CF is subject to the following equality constraint [36]:

$$y_k=w^{T}\varphi(x_k)+b+e_k \tag{14}$$

where γ is a tuning parameter in LSSVM which is used to prevent overfitting, and e_k is the variable error for training data. The Lagrangian form of LSSVM takes the following form [36]:

$$L(w, b, e, a)=\frac{1}{2}w^{T}w+\frac{1}{2}\gamma\sum_{k=1}^{N}e_k^2-\sum_{k=1}^{N}a_k(w^{T}\varphi(x_k)+b+e_k-y_k) \tag{15}$$

where a_k are Lagrange multipliers. To find the solution, the derivatives of Eq. (15) should be equated to zero, which results in the following relations [36]:

$$\begin{cases} \dfrac{\partial L}{\partial w}=0\rightarrow w=\sum_{k=1}^{N}a_k\varphi(x_k) \\ \dfrac{\partial L}{\partial b}=0\rightarrow\sum_{k=1}^{N}a_k=0 \\ \dfrac{\partial L}{\partial e_k}=0\rightarrow a_k=\gamma e_k,\ k=1, 2, 3, \ldots, N \\ \dfrac{\partial L}{\partial e_k}=0\rightarrow w^{T}\varphi(x_k)+b+e_k-y_k=0,\ k=1, 2, 3, \ldots, N \end{cases} \tag{16}$$

This system of equations should be solved to find parameters of LSSVM. In this study the radial basis function (RBF) kernel was applied for both SVM and LSSVM models, because this function has good performance for nonlinear systems [36]:

$$K(x, x_k)=\exp\left(\frac{-\|x_k-x\|^2}{\sigma^2}\right) \tag{17}$$

where σ^2 is an adjustable parameter. Therefore, LSSVM has two adjustable parameters of σ^2 and γ which should be determined by minimizing an objective function such as mean square deviation (MSD):

In this study, we used genetic algorithm (GA) [38] to optimize the parameters of LSSVM. A schematic diagram of the applied GA-LSSVM is illustrated in Fig. 3. According to this figure, input data are divided into training and testing set. In addition, GA opti-

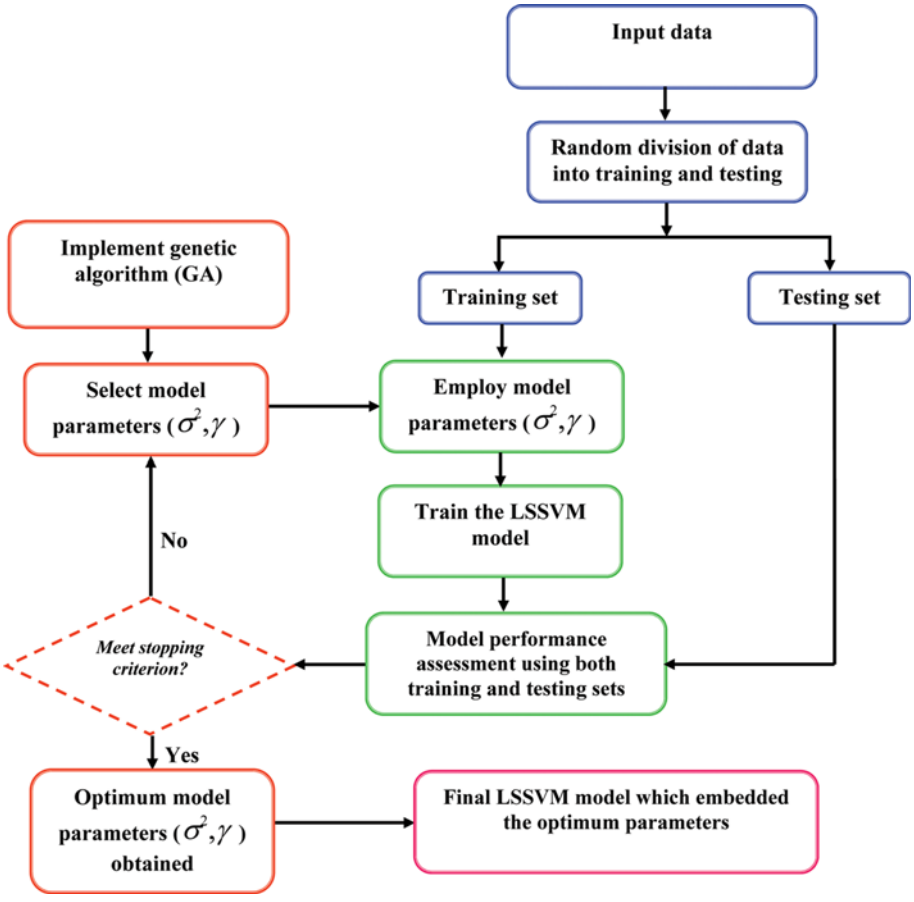

Fig. 3. Schematic of procedure for development of GA-LSSVM model.

mizes model parameters based on training data set. Then, the performance of the model is evaluated by comparing estimated data by LSSVM model with the train and test set. If the model meets the stopping criterion for both training and test sets, the model is saved, otherwise new parameters will be produced according to the GA algorithm and again the LSSVM model will be trained until meeting the stopping criterion.

4. Genetic Algorithm

The GA technique, as one kind of evolutionary algorithm, mimics natural evolution to find global optima in complex problems [39]. In the GA approach, biological methods comprising mutation, selection, crossover and inheritance are used to optimize problems. The GA technique finds optimum parameters of LSSVM model in this study. The process in GA approach is divided into five steps which are described as follows.

4-1. Initial Population

To begin the evolution, an initial population of chromosomes (individuals) should be randomly generated [40]. Initializing leads to creating an initial population as $[\sigma, \gamma]_n$ in which n stands for the number of generation. Iterative search is started implementing the initial population as a start point. Note that population size, mutation probability, crossover probability and selection probability are determined according to the problem under study.

4-2. Fitness Function

The fitness function has a key role for evaluation of a chromosome. In this study, the fitness function is MSD between estimated data and experimental data. Thus, generated chromosomes, which are parameters of the LSSVM model, are employed to LSSVM, and MSD is calculated as follows:

$$\mathrm{MSD}=\frac{1}{\mathrm{N}}\sum_{i=1}^{N}(\mathrm{P}^{calc,\,i}-\mathrm{P}^{exp,\,i})^2 \tag{18}$$

Then, the fitness value of a chromosome is calculated as follows:

$$\mathrm{F}_{fitness}=\frac{1}{1+\mathrm{MSD}} \tag{19}$$

Consequently, a smaller value of MSD for GA-LSSVM leads to closer fitness value to 1. After calculation of fitness magnitude of all chromosomes, based on the generic operators of GA approach (selection, crossover and mutation) a population of chromosomes is updated.

4-3. Selection Operation

This operation is used to choose the chromosomes with large fitness in the population. In this study, the roulette wheel selection approach was implemented to choose new chromosomes.

4-4. Crossover Operation

Using this operation, it is possible to create variation in the population. This is performed such that an offspring in which a new chromosome is created from its parent's chromosome. In this procedure, some genes of two chromosomes of parents are replaced to create a new chromosome.

4-5. Mutation Operation

To maintain genetic diversity from one generation of a population of chromosomes to the next one, this operator is implemented in GA. By the mutation, the gene value of chromosomes is changed to create a new one. Thus, it is possible to obtain entirely different solution by using mutation. The probability of mutation has a key role in GA and should be set to a low value. Consideration of high value for probability leads to turning of GA to a primitive random search.

Good and detailed information of GA can be find in the literature [38,40].

5. Group Method of Data Handling

As the first step, a nonhomogeneous matrix is generated by algorithm [30,31]:

$$\begin{matrix} O_1 & I_{11} & I_{12} & \cdots & I_{1m} \\ O_2 & I_{21} & I_{22} & \cdots & I_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ O_n & I_{n1} & I_{n2} & \cdots & I_{nm} \end{matrix} \quad (20)$$

where, n is the number of experimental data points and m represents the number of independent parameters. In the next step, independent parameters are substituted with new parameters called "virtual parameters" and are demonstrated with V here:

$$V_{ij}=a+b_1I_i+b_2I_j+c_1I_i^2+c_2I_j^2+dI_iI_j \quad j=1, 2, \ldots, M \quad (21)$$

As can be seen, only two independent parameters at a time are considered to develop the virtual parameters. A least square approach is applied to determine the coefficients in Eq. (19) with minimization of the following objective function for each column:

$$F_j^2=\sum_{i=1}^{n_t}(O_i-V_{ij})^2 \quad j=1, 2, \ldots, P \quad (22)$$

where n_t is the number of data points for training procedure. Note that an arbitrary ratio is randomly chosen to divide the experimental data into training set and testing set. The training data points are applied to construct polynomials, in order to find the most precise ones.

The final mathematical structure of the GMDH model is an explicit analytical model including input layer, intermediate layers, and output layer in which several nodes combine interaction of virtual and independent parameters.

Two disadvantages are generally related to the GMDH model: considering only two independent parameters at a time, and also connecting of parameters in a layer with only the next layer [41]. Such limitations lead to low flexibility of the model and consequently poor ability for modeling non-linear systems. In this regard, the hybrid GMDH-type polynomial neural network (GMDH-PNN) can be good alternative for GMDH model, which has not the mentioned drawbacks. The grand polynomial equation of GMDH-PNN has the following form:

$$O=a+\sum_{i=1}^{M}\sum_{j=1}^{M}\cdots\sum_{k=1}^{M}b_{ij\cdots k}I_i^nI_j^n\cdots I_k^n \quad n=0, 1, \ldots, 2^l \quad (23)$$

where l represents the number of layers.

Table 5. Tunable parameters of GA-LSSVM and GA-SVM models

Property	GA-LSSVM parameters		GA-SVM parameters		
	γ	σ^2	ε	c	σ^2
Density	724	0.2704	0.013	189	0.3136
Heat capacity	1121	0.6084	0.003	985	0.5776
Thermal conductivity	448	1.7635	0.012	224	1.8769
Surface tension (water based)	1232	0.1764	0.010	420	0.2704
Surface tension (organic based)	931	0.8100	0.015	122	0.5929

Table 6. Statistical parameters of the models for different properties

Property	Model	MARD%	MD%	RMSD%	R^2
Density	GA-LSSVM	0.3886	−0.0097	0.0742	0.9999
	GA-SVM	0.6218	−0.0964	0.1077	0.9981
	GMDH-PNN	1.0704	0.0103	0.1694	0.9994
Heat capacity	GA-LSSVM	0.1802	−0.6202	6.3050	0.9999
	GA-SVM	0.3212	−3.7711	8.8505	0.9998
	GMDH-PNN	0.8449	1.1538	24.9169	0.9996
Thermal conductivity	GA-LSSVM	0.7758	0.0805	0.0412	0.9893
	GA-SVM	1.0392	0.0575	0.0332	0.9928
	GMDH-PNN	2.2561	−0.2761	0.1288	0.9865
Surface tension	GA-LSSVM	1.1060	0.4538	8.0845	0.9982
	GA-SVM	1.2177	−1.8272	9.0293	0.9989
	GMDH-PNN	2.9410	1.6283	18.1177	0.9949

GMDH-PNN models are more flexible and have more ability to trace the trend of experimental data for non-linear systems [2,23, 33,34]. For this reason, in the present study the GMDH-PNN was applied for modeling physicochemical properties of IL mixtures.

RESULTS AND DISCUSSION

To develop GA-LSSVM, GA-SVM, and GMDH-PNN for the mentioned properties of this work, first the experimental data was randomly divided into two sets of training and testing data. 60% of the experimental data was randomly devoted to train set and the remaining 40% was selected as test data. It was also possible to consider higher number of data points for training set like 70/30 or 80/20 training/test split. However, for a reliable model it is more interesting to consider higher number of data points for the test-

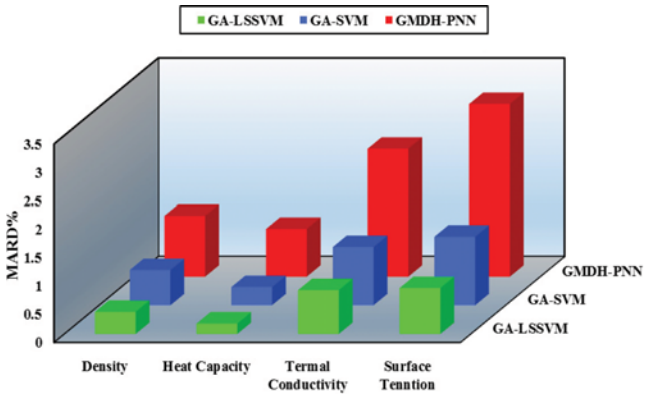

Fig. 4. MARD% of proposed models for physicochemical properties of ILs mixtures.

Table 7. MARD% of the models for estimation of density

No.	Binary system	MARD%		
		GA-LSSVM	GA-SVM	GMDH-PNN
1	[C4MIm][PF6] (1)+[C4MIm][BF4] (2)	0.99	0.31	1.56
2	[C6MIm][BF4] (1)+[C2MIm][BF4] (2)	0.49	0.24	4.06
3	[C6MIm][Cl] (1)+[C6MIm][PF6] (2)	1.27	1.42	5
4	[C8MIm][Cl] (1)+[C8MIm][BF4] (2)	0.91	1.18	2.09
5	2-Butanol (1)+[MOA][BTI] (2)	0.28	0.35	1.22
6	2-Butanone (1)+[C4MIm][PF6] (2)	0.52	0.46	0.89
7	2-Butanone (1)+[Mmim][MeSO4] (2)	0.2	0.5	0.82
8	2-Propanol (1)+[MOA][BTI] (2)	0.59	0.99	2.33
9	Acetic acid (1)+[BMIM][SCN] (2)	0.18	0.28	0.44
10	Acetic acid (1)+[EMIM][EtSO4] (2)	0.51	0.38	0.4
11	Acetone (1)+[C4MIm][PF6] (2)	0.73	0.62	1.08
12	Acetonitrile (1)+[BMIM][SCN] (2)	0.61	0.52	1.04
13	Acetonitrile (1)+[EMIM][EtSO4] (2)	0.52	0.33	1.17
14	Dimethyl carbonate (1)+[C6MIm][PF6] (2)	1.44	1.31	0.46
15	Ethanol (1)+[C4MIm][BF4] (2)	1.56	2.28	0.76
16	Ethanol (1)+[Mmim][MeSO4] (2)	0.42	0.34	0.37
17	Ethyl acetate (1)+[C4MIm][PF6] (2)	0.3	0.28	1.18
18	Ethyl acetate (1)+[Mmim][MeSO4] (2)	0.19	0.44	0.68
19	Methanol (1)+[C4MIm][SCN] (2)	0.57	0.46	1.02
20	Methyl acetate (1)+[C4MIm][BF4] (2)	0.42	0.46	1.3
21	Methyl formate (1)+[C4MIm][BF4] (2)	0.57	0.23	0.41
22	n-Hexane (1)+[C8MIm][PF6] (2)	0.77	0.71	0.65
23	Propionic acid (1)+[BMIM][SCN] (2)	0.19	0.29	0.62
24	Propionic acid (1)+[EMIM][EtSO4] (2)	0.81	0.73	1.01
25	Water (1)+[C2MIm][BTI] (2)	0.15	0.12	1.17
26	Water (1)+[C4MPyr][BTI] (2)	0.08	0.31	0.72
27	Water (1)+[C6MIm][BF4] (2)	0.25	0.29	0.97
28	Water (1)+[C6MIm][Cl] (2)	0.49	0.49	0.41
29	Water (1)+[C8MIm][Cl] (2)	0.23	0.31	0.89
30	Water (1)+[empy][EtSO4] (2)	0.99	1.25	1.95
31	Water (1)+[N1114][BTI] (2)	0.08	0.2	0.69
32	Water (1)+[pDMIM][BF4] (2)	0.61	0.89	1.23
33	Water (1)+[C4(3-m)Py][BF4] (2)	0.39	0.63	1.17
Overall		0.38	0.62	1.07

ing set as is possible. Adjustable parameters of GA-LSSVM and GA-SVM are presented in Table 5. The optimized GMDH-PNN models for all properties are reported in Tables S1 to S5 in the supplementary material file.

Statistical parameters of different models for the prediction of physicochemical properties of IL mixtures are summarized in Table 6. In addition, MARD% of the proposed models are depicted in Fig. 4 for all the properties. As can be seen, the GMDH-PNN model has lower accuracy than other models. Also, GA-LSSVM is more accurate than GA-SVM for the prediction of all physicochemical properties. In the next sections, the performance of the developed models is investigated with more details for the aforementioned physicochemical properties.

1. Density

Table 7 shows MARD% of the proposed models for estimation of density of IL mixtures. Overall MARD% for all systems and for GA-LSSVM, GA-SVM and GMDH-PNN are 0.38, 0.62 and 1.07, respectively. Thus, SVM based models are more accurate than GMDH-PNN, and GA-LSSVM provided slightly more accurate results.

To better visualize this issue and understand the performance of the models, it would be worthwhile to investigate the cumulative frequency of data points as a function of absolute relative error for the models, which is portrayed in Fig. 5. This figure also confirms the better performance of GA-LSSVM and GA-SVM models in comparison to GMDH-PNN. Moreover, as more detail, for GA-LSSVM, 53% of data points have absolute relative error less than 0.25%, while for GA-SVM 35% of data points have such error. Altogether GA-LSSVM is more accurate than other models; hence, performance of this model for all density data points is more discussed here.

Fig. 6 illustrates the experimental data and estimated values of density by GA-LSSVM for both train and test sets versus data indices. The ability of GA-LSSVM is noticeably confirmed due to close match between the estimated data and the experimental density data points.

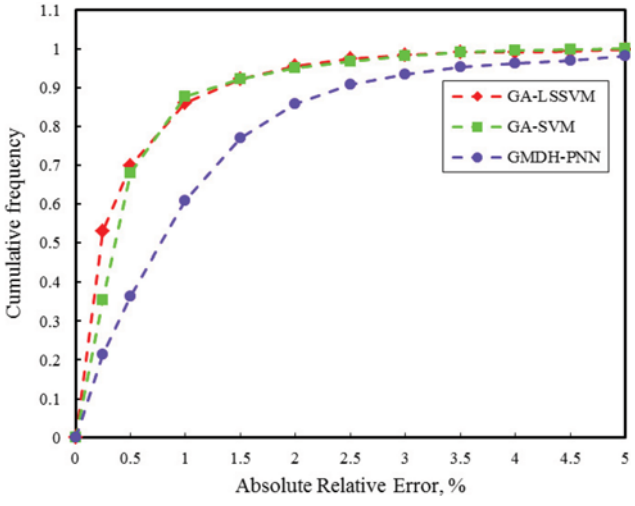

Fig. 5. Cumulative frequency of the proposed models for prediction of density as a function of absolute relative error.

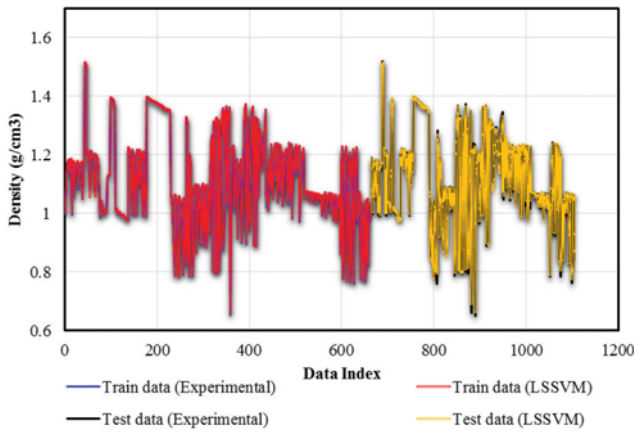

Fig. 6. Comparison between the predicted values by GA-LSSVM and experimental density data versus data index.

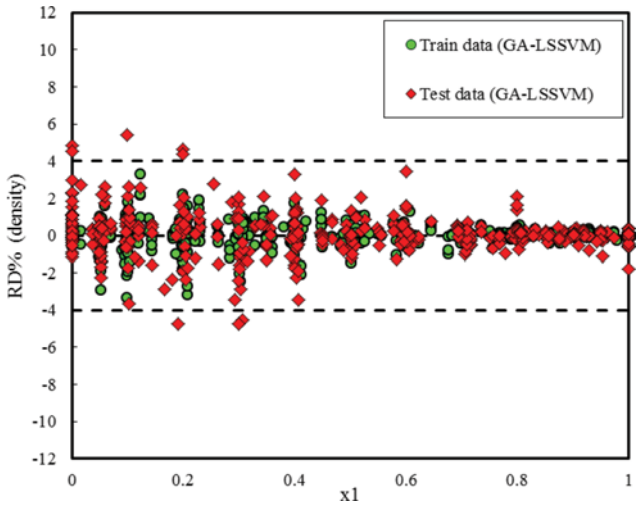

Fig. 7. RD% of data points for GA-LSSVM versus for prediction of density.

Fig. 7 demonstrates relative deviation (RD%) of data points as a function of mole fraction of IL in mixture (x_1). As seen, RD% of most data points lie between, demonstrating good accuracy of GA-LSSVM model. This figure well reveals the applicability of GA-LSSVM with variation of x_1. As x_1 increases, the accuracy of GA-LSSVM model for predicting density of mixture is reduced. Then, the model is more reliable for mixtures with higher mole fractions of IL in mixture. However, as already mentioned, because RD% of most of the data points lies between ±4, the proposed GA-LSSVM model has enough accuracy for calculations even at low x_1 values.

2. Heat Capacity

Table 8 reports MARD% of the proposed models for estimation of heat capacity of IL mixtures. As can be seen, the proposed models have good accuracy for estimation of heat capacity of different mixtures. Furthermore, SVM based models are more accurate than GMDH-PNN. Also, Overall MARD% of GA-LSSVM is slightly lower than GA-SVM. The maximum MARD% of GA-LSSVM, GA-SVM and GMDH-PNN are 3.24, 3.47, and 5.83 respectively,

Table 8. MARD% of the models for estimation of heat capacity

No.	Binary system	MARD%		
		GA-LSSVM	GA-SVM	GMDH-PNN
1	Acetonitrile (1)+[C6MIm][BF4] (2)	0.14	0.35	0.54
2	Acetonitrile (1)+[C8MIm][BF4] (2)	0.17	0.3	1.14
3	Ethanol (1)+[bmpyr][BF4] (2)	0.16	0.31	1.15
4	Methanol (1)+[C6MIm][BF4] (2)	0.18	0.37	0.97
5	Methanol (1)+[C8MIm][BF4] (2)	0.23	0.36	1.02
6	Water (1)+[C4MIm][BF4] (2)	0.15	0.23	0.83
7	Water (1)+[C4MIm][MeSO4] (2)	0.23	0.27	0.77
8	Water (1)+[C4MIm][PF6] (2)	0.09	0.19	0.7
9	Water (1)+[C4MIm][TfO] (2)	0.25	0.38	0.92
Overall		0.18	0.32	0.84

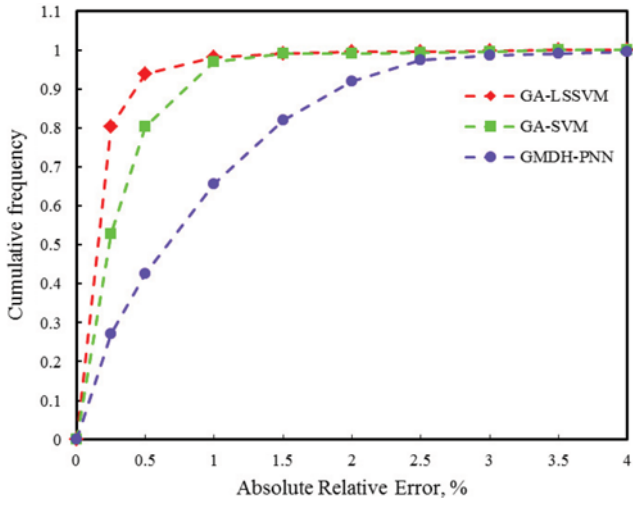

Fig. 8. Cumulative frequency plot of the proposed models for prediction of heat capacity as a function of absolute relative error.

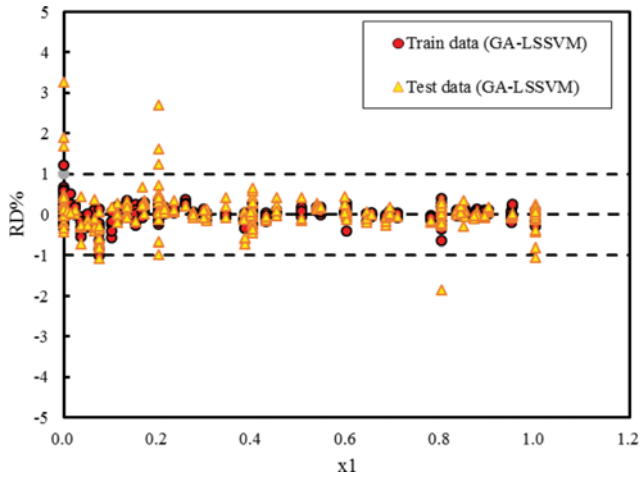

Fig. 9. RD% of the data points for GA-LSSVM versus for prediction of heat capacity.

demonstrating high accuracy of the proposed models.

Fig. 8 displays cumulative frequency of the models as a function of absolute relative error. For GA-LSSVM model, 80% of the data points have absolute relative error lower than 0.25, while for GA-SVM model 52% of data points have such error. This situation for GMDH-PNN model is only 27%, which confirms the lower accuracy of this model compared to the other models.

Fig. 9 illustrates RD% of data points as a function of mole fraction of IL (x_1) for GA-LSSVM model. As can be observed, RD% of most of the data points lies between ±1, indicating excellent accuracy of GA-LSSVM model for estimation of heat capacity of IL mixtures.

3. Thermal Conductivity

Despite the importance of thermal properties of ILs, such as thermal conductivity for safe and efficient design of heat transfer equipment, few experimental investigations have been performed on this area, both for pure ILs and mixtures containing ILs [42]. However, we have used limited available data for validation of the proposed models. Table 9 represents MARD% of the proposed models for IL mixtures. GA-LSSVM exhibits better accuracy than GA-SVM model for prediction of thermal conductivity of mixtures.

Table 9. MARD% of the models for estimation of thermal conductivity

No.	Binary system	MARD%		
		GA-LSSVM	GA-SVM	GMDH-PNN
1	Water (1)+[C2mim][EtSO4] (2)	1.02	1.07	0.89
2	Methanol (1)+[Mmim][DMP] (2)	0.16	0.99	4.04
3	Water (1)+[C4MIm][TfO] (2)	1.34	1.04	1.83
Overall		0.77	1.03	2.25

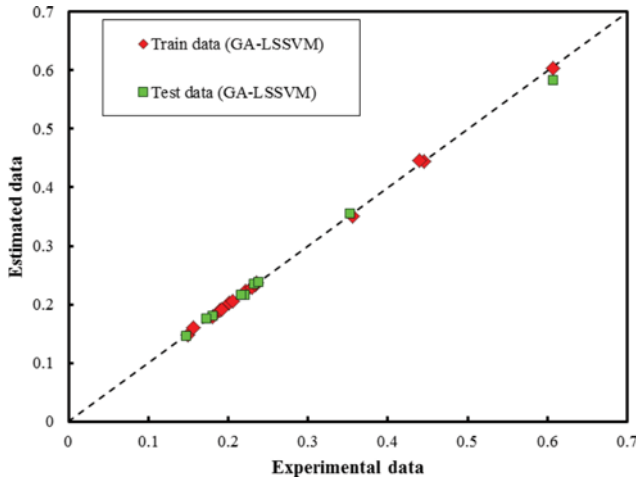

Fig. 10. Comparison between estimated values by GA-LSSVM and experimental thermal conductivity data.

In addition, a comparison between estimated thermal conductivity data by GA-LSSVM and experimental data for both train and test data is shown in Fig. 10. As seen, data points keep close to the diagonal line, indicating excellent conformity between the experimental data and the results obtained by the proposed GA-LSSVM model.

4. Surface Tension

Compared to the other investigated properties of IL mixtures in previous sections, surface tension has more complex behavior. For instance, for the surface tension behavior of water-based and organic-based IL mixtures, opposite trends have been reported in experimental investigations [7]. It must be stressed that the surface tension of water is larger than most of ILs at 298 K, while most organic compounds have lower ones. Thus, because of lower surface tension of ILs than water, generally surface tension of water-based IL mixtures exhibits descending behavior with increasing mole fraction of IL. While, for the case of organic-based an ascending behavior is observed. Excellent descriptions and more detail information for this phenomenon have been reported by Tariq et al. [43]. Therefore, two different models for organic based mixtures and water based mixtures have been developed. Table 10 shows the results of the proposed models for estimation of water based IL mixtures. Overall MARD% of GA-LSSVM, GA-SVM is approximately equal (with slightly better accuracy for GA-LSVM), which demonstrates the same performance for both of the models. However, GMDH-PNN is not accurate in comparison to GA-LSSVM and GA-SVM. This shows that flexibility of GMDH-PNN for modeling complex and nonlinear systems is lower than SVM based models, which is due to considering simple polynomials as transfer functions for GMDH-PNN based models.

Table 11 demonstrates MARD% of different models for estimation of surface tension for organic based mixtures. As illustrated, GA-LSSVM has slightly better results than GA-SVM. Overall, MARD% of GA-LSSVM and GA-SVM are 0.92 and 1.02, respectively. Although from the accuracy point of view GA-LSSVM has not a significant excellence compared to GA-SVM, the advantage of GA-LSSVM is its high speed of convergence. Therefore, GA-LSSVM is yet more appropriate for estimation of surface tension of IL mixtures. Apart from that, regarding GMDH-PNN the overall MARD% is 3.55 which is a reasonable value for estimation of surface tension of organic based IL mixtures.

With a plot of cumulative frequency versus absolute relative error, the same accuracy of GA-LSSVM and GA-SVM and also their superiority over GMDH-PNN model have been confirmed in Fig. 11(a) (water based) and Fig. 11(b) (organic based). As already stated, with considering high convergence speed of GA-LSSVM, it is the more preferable model.

Fig. 12 shows a plot of estimated values by GA-LSSVM versus experimental surface tension data for both organic-based and water-based mixtures. As depicted, the data points for organic-based mixtures accumulate close to the diagonal line, indicating good conformity between the experimental data and the results obtained by the proposed GA-LSSVM model. Besides, there is a reasonable con-

Table 10. MARD% of the models for estimation of surface tension for water based mixtures

No.	Binary system	MARD%		
		GA-LSSVM	GA-SVM	GMDH-PNN
1	Water (1)+[C2MIm][BF4] (2)	1.67	3.39	6.45
2	Water (1)+[C1MIm][MeSO4] (2)	1.49	0.49	7.07
3	Water (1)+[C2MIm][C4SO4] (2)	1.02	0.96	2.26
4	Water (1)+[C2MIm][EtSO4] (2)	1.85	2	3.14
5	Water (1)+[C2MIm][MeSO3] (2)	1.84	1.66	2.93
6	Water (1)+[C4MIm][BF4] (2)	2.05	1.82	5.21
7	Water (1)+[C4MIm][Gly] (2)	0.18	0.54	1.49
8	Water (1)+[C4Py][NO3] (2)	1.91	1.54	4.44
9	Water (1)+[C6MIm][BF4] (2)	3.67	5.94	1.48
10	Water (1)+[N112(hoe)][Br] (2)	1.21	1.34	5.95
11	Water (1)+[N311(hoe)][Br] (2)	8.34	4.98	22.99
12	Water (1)+[P666(14)][BTI] (2)	4.12	3.93	5.46
13	Water (1)+[P666(14)][Dca] (2)	1.53	2.29	2.36
Overall		1.45	1.58	3.41

Table 11. MARD% of the models for estimation of surface tension for organic based mixtures

No.	Binary system	MARD%		
		GA-LSSVM	GA-SVM	GMDH-PNN
1	Methanol (1)+[C1MIm][MeSO4] (2)	0.56	0.71	2.20
2	1-butanol (1)+[C4MIm][BTI] (2)	3.27	2.45	8.22
3	1-propanol (1)+[C4MIm][BTI] (2)	0.94	1.56	3.40
4	Acetonitrile (1)+[C2MIm][BTI] (2)	0.50	0.85	2.76
5	Acetonitrile (1)+[C4MIm][BTI] (2)	0.56	0.91	2.60
6	Dimethyl sulfoxide (1)+[C2MIm][BTI] (2)	0.37	0.68	1.74
7	Dimethyl sulfoxide (1)+[C4MIm][BTI] (2)	0.58	0.92	2.45
8	Ethanol (1)+[C2MIm][C6SO4] (2)	1.61	1.43	2.76
9	Ethanol (1)+[C2MIm][C8SO4] (2)	1.36	1.09	3.30
10	Ethanol (1)+[C4MIm][BF4] (2)	2.57	0.88	7.87
11	Ethanol (1)+[C6MIm][BF4] (2)	1.12	0.55	3.10
12	Ethanol (1)+[C8MIm][BF4] (2)	3.54	3.48	1.35
13	Methanol (1)+[C2MIm][MeSO4] (2)	1.90	1.32	9.33
14	Tetrahydrofuran (1)+[C2MIm][BTI] (2)	0.50	0.83	0.93
15	Tetrahydrofuran (1)+[C4MIm][BTI] (2)	0.82	0.83	1.30
Overall		0.92	1.02	2.68

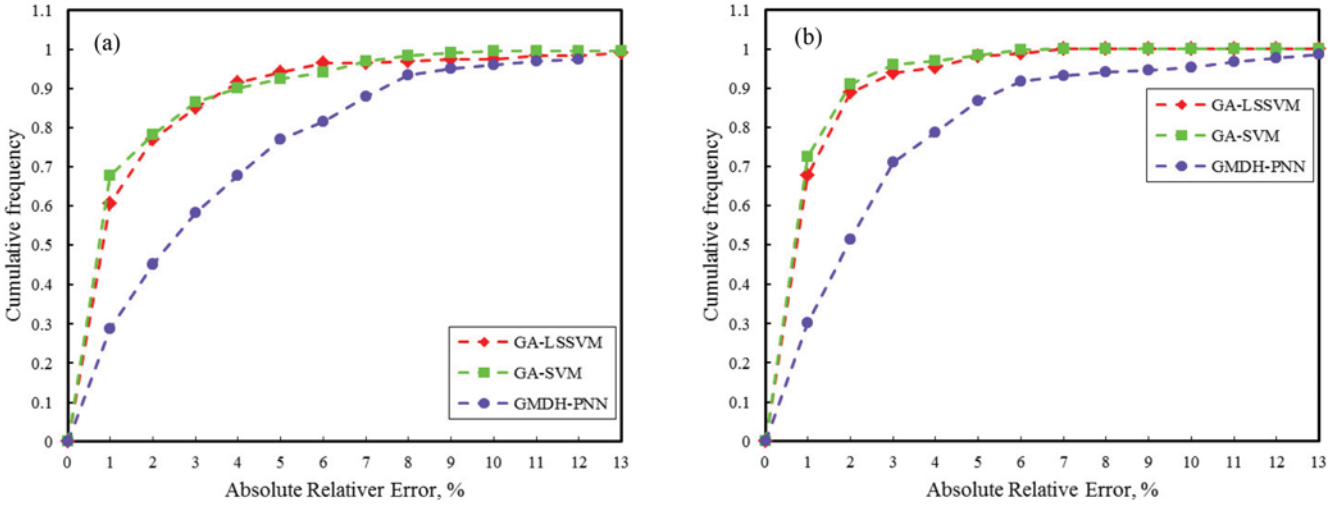

Fig. 11. Cumulative frequency plot of the proposed models for prediction of surface tension as a function of absolute relative error: (a) Water based, (b) organic based.

formity between estimated data by the GA-LSSVM and experimental surface tension data of water-based mixtures.

Fig. 13 exhibits relative deviation of GA-LSSVM for train and test data as a function of mole fraction of ionic liquids for both organic-based and water-based mixtures. For all studied systems, most of the data points lie between the lines demonstrative relative deviation of ±6% which is a reasonable performance for prediction of surface tension of IL mixtures.

As described, the surface tension behavior of the organic-based and water-based IL mixtures exhibits opposite trends. The mixture of water-[C4MIm][Gly] is considered as a water-based IL mixture. The experimental surface tension data and the predicted results of GA-LSSVM model for different temperatures as a function of mole fraction of IL are shown in Fig. 14. The surface tension of the mixture is reduced with increasing mole fraction of ionic liquid, and as seen, there is a reasonable conformity between the predicted results of GA-LLSVM and the experimental data. Furthermore, GA-LSSVM can follow the physically expected trend of variation of surface tension with temperature in the mixture.

The mixture of acetonitrile-[C2MIm] is considered as an organic based mixture. Experimental surface tension data and the predicted values by GA-LSSVM as a function of mole fraction of IL are shown in Fig. 15. In contrast to the water-[C4MIm][Gly] mixture, for acetonitrile-[C2MIm] mixture surface tension is enhanced with increasing mole fraction of IL. As shown, GA-LSSVM can predict the trend of experimental data with a reasonable performance.

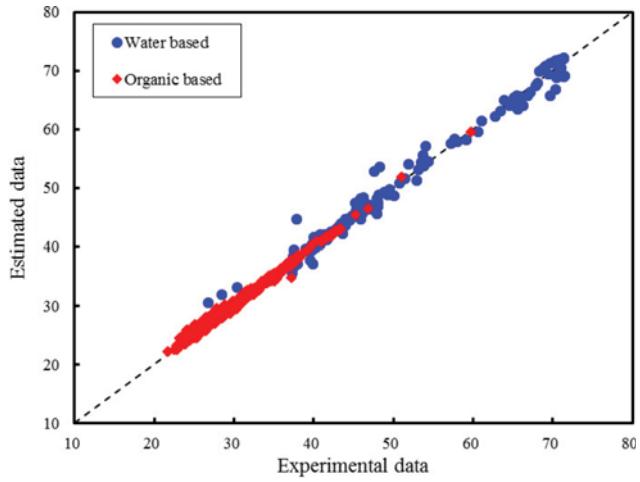

Fig. 12. Comparison between estimated values by GA-LSSVM and experimental surface tension data for both organic-based and water-based mixtures.

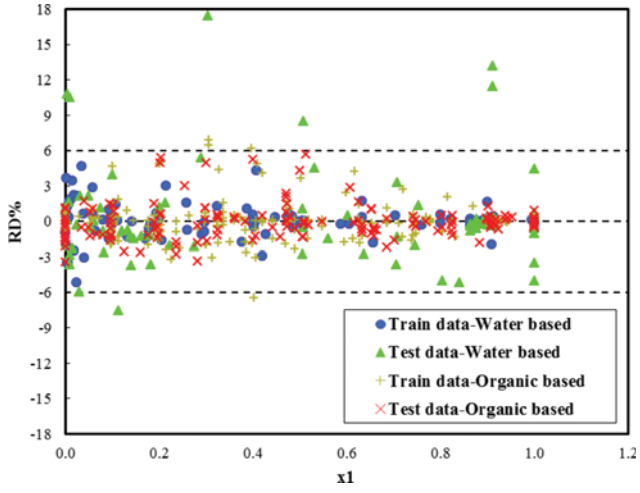

Fig. 13. RD% of the data points for GA-LSSVM versus for prediction of surface tension.

CONCLUSION

The present study showed that GA-LSSVM is a robust and accurate model for prediction of physicochemical properties of IL mixtures, including density, heat capacity, thermal conductivity, and surface tension. MARD% of GA-LSSVM for density, heat capacity, thermal conductivity and surface tension are 0.38%, 0.18%, 0.77% and 1.18%, respectively. Although GA-LSSVM and GA-SVM has approximately the same level of accuracy, GA-LSSVM has simpler mathematical development, which leads to higher speed of convergence, and consequently it is computationally less expensive. The results of this study demonstrated that GA-LSSVM can be an appropriate candidate for prediction of physicochemical properties of IL mixtures. On the other hand, it was clearly demonstrated that the GMDH-PNN model has lower accuracy compared to GA-LSSVM. However, GMDH-PNN still can be a useful model with simple use for prediction of physicochemical properties of IL mixtures, especially in engineering applications. In summary, the successful development of predictive models for the determination of some physicochemical properties of IL mixtures in the present study allows us to promote these methods for other physicochemical properties of IL mixtures in future studies to overcome the limitations for development of IL based industries and equipment.

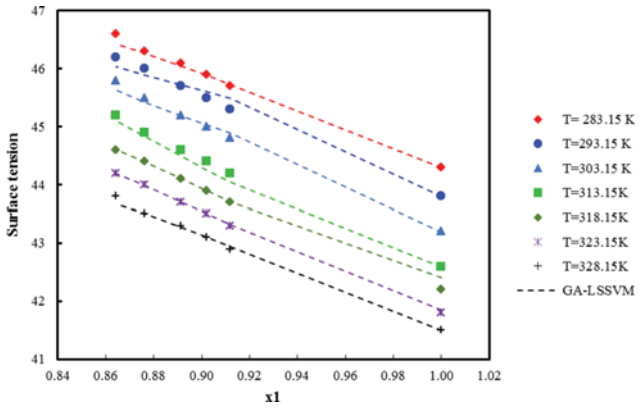

Fig. 14. Experimental surface tension data and predicted results of GA-LSSVM model for water-[C4MIm][Gly] mixture.

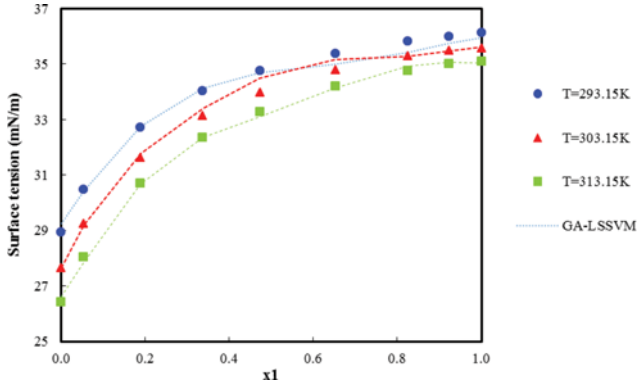

Fig. 15. Experimental surface tension data and predicted results of GA-LSSVM model for acetonitrile-[C2MIm][BTI].

ABBREVIATIONS

IL : ionic liquid
[C2MIm][BF4] : 1-ethyl-3-methylimidazolium tetrafluoroborate
[C4MIm][BF4] : 1-butyl-3-methylimidazolium tetrafluoroborate
[C6MIm][BF4] : 1-hexyl-3-methylimidazolium tetrafluoroborate
[C1MIm][MeSO4] : 1,3-dimethylimidazolium methylsulfate
[C2MIm][MeSO3] : 1-ethyl-3-methyl imidazolium methane sulfonate
[C2MIm][EtSO4] : 1-ethyl-3-methylimidazolium ethylsulfate
[C2MIm][C4SO4] : 1-ethyl-3-methylimidazolium n-butylsulfate
[C4MIm][Gly] : 1-butyl-3-methylimidazolium aminoacetate
[C4Py][NO3] : 1-butylpyridinium nitrate
[N311(hoe)][Br] : propyl(2-hydroxyethyl)dimethylammonium bromide
[N112(hoe)][Br] : ethyl(2-hydroxyethyl)dimethylammonium bromide

[P666(14)][Dca] : trihexyl(tetradecyl)phosphonium dicyanamide
[P666(14)][BTI] : trihexyltetradecylphosphonium bis(trifluoromethylsulfonyl)imide
[C1MIm][MeSO4] : 1,3-dimethylimidazolium methylsulfate
[C2MIm][MeSO4] : 1-ethyl-3-methylimidazolium methylsulfate
[C4MIm][BF4] : 1-butyl-3-methylimidazolium tetrafluoroborate
[C6MIm][BF4] : 1-hexyl-3-methylimidazolium tetrafluoroborate
[C8MIm][BF4] : 1-octyl-3-methylimidazolium tetrafluoroborate
[C2MIm][C6SO4] : 1-ethyl-3-methylimidazolium n-hexylsulfate
[C2MIm][C8SO4] : 1-ethyl-3-methylimidazolium octyl sulfate
[C4MIm][BTI] : 1-butyl-3-methylimidazolium bis((trifluoromethyl) sulfonyl)imide
[C2MIm][BTI] : 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide
[C_4MIm][PF_6] : 1-butyl-3-methylimidazolium hexafluorophosphate
[C4MIm][TfO] : 1-butyl-3-methylimidazolium trifluoromethylsulfonate
[C4MIm][MeSO4] : 1-butyl-3-methylimidazolium methylsulfate
[C6MIm][BF_4] : 1-hexyl-3-methylimidazolium tetrafluoroborate
[C8MIm][BF4] : 1-octyl-3-methylimidazolium tetrafluoroborate
[bmpyr][BF_4] : 1-butyl-3-methylpyridinium tetrafluoroborate
[C_2mim][$EtSO_4$] : 1-ethyl-3-methylimidazolium ethylsulfate
[mmim]DMP : N-methyl-N-methylimidazolium dimethyl phosphate
[C4(3-m)Py][BF4] : 1-butyl-3-methylpyridinium tetrafluoroborate
[empy][EtSO4] : 1-ethyl-3-methylpyridinum ethylsolfate
[C6MIm][Cl] : 1-hexyl-3-methylimidazolium chloride
[C4MPyr][BTI] : 1-butyl-3-methylpyrrolidinium bis(trifluoromethylsulfonyl)imide
[C8MIm][Cl] : 1-hexyl-3-methylimidazolium chloride
[pDMIM][BF4] : 1-propyl-2,3-dimethylimidazolium tetrafluoroborate
[N1114][BTI] : butyltrimethylammonium bis(trifluoromethylsulfonyl)imide
[C4MIm][SCN] : 1-butyl-3-methylimidazolium thiocyanate
[Mmim][$MeSO_4$] : 1,3-dimethylimidazolium methyl sulphate
[MOA][BTI] : methyl trioctylammonium bis(trifluoromethylsulfonyl)imide
[C8MIm][PF6] : 1-octyl-3-methylimidazolium hexafluorophosphate
[C6MIm][PF6] : 1-hexyl-3-methylimidazolium hexafluorophosphate
[C2MIm][BF4] : 1-ethyl-3-methylimidazolium tetrafluoroborate

SUPPORTING INFORMATION

Additional information as noted in the text. This information is available via the Internet at http://www.springer.com/chemistry/journal/11814.

REFERENCES

1. B. Mokhtarani, M. M. Mojtahedi, H. R. Mortaheb, M. Mafi, F. Yazdani and F. Sadeghian, *J. Chem. Eng. Data*, **53**, 677 (2008).
2. S. Atashrouz, M. Mozaffarian and G. Pazuki, *Ind. Eng. Chem. Res.*, **54**, 8600 (2015).
3. B. H. Park, *Korean J. Chem. Eng.*, **33**, 2191 (2016).
4. B. Gonzalez, N. Calvar, E. Gomez, I. Dominguez and A. Dominguez, *J. Chem. Eng. Data*, **54**, 1353 (2009).
5. T.-Y. Wu, B.-K. Chen, L. Hao, K.-F. Lin and I.-W. Sun, *J. Taiwan Inst. Chem. Eng.*, **42**, 914 (2011).
6. S. Atashrouz, M. Zarghampour, S. Abdolrahimi, G. Pazuki and B. Nasernejad, *J. Chem. Eng. Data*, **59**, 3691 (2014).
7. Y. Huang, X. Zhang, Y. Zhao, S. Zeng, H. Dong and S. Zhang, *Phys. Chem. Chem. Phys.*, **17**, 26918 (2015).
8. K. Marsh, J. Boxall and R. Lichtenthaler, *Fluid Phase Equilib.*, **219**, 93 (2004).
9. A. Heintz, *J. Chem. Thermodyn.*, **37**, 525 (2005).
10. H. Niedermeyer, J. P. Hallett, I. J. Villar-Garcia, P. A. Hunt and T. Welton, *Chem. Soc. Rev.*, **41**, 7780 (2012).
11. V. H. Alvarez, S. Mattedi, M. Martin-Pastor, M. Aznar and M. Iglesias, *J. Chem. Thermodyn.*, **43**, 997 (2011).
12. S. M. Hosseini, J. Moghadasi, M. M. Papari and F. Fadaei Nobandegani, *Ind. Eng. Chem. Res.*, **51**, 758 (2012).
13. F. Yousefi, *Ionics (Kiel)*, **18**, 769 (2012).
14. A. E. Andreatta, E. Rodil, A. Arce and A. Soto, *J. Solution Chem.*, **43**, 404 (2014).
15. P.-Y. Lin, A. N. Soriano, R. B. Leron and M.-H. Li, *Exp. Therm. Fluid Sci.*, **35**, 1107 (2011).
16. E. J. Gonzalez, B. Gonzalez, N. Calvar and A. Dominguez, *J. Chem. Eng. Data*, **52**, 1641 (2007).
17. B. González, N. Calvar, E. Gómez and Á. Domínguez, *J. Chem. Thermodyn.*, **40**, 1274 (2008).
18. E. Mohagheghian, H. Zafarian-Rigaki, Y. Motamedi-Ghahfarrokhi and A. Hemmati-Sarapardeh, *Korean J. Chem. Eng.*, **32**, 2087 (2015).
19. A. A. Babaei, A. Khataee, E. Ahmadpour, M. Sheydaei, B. Kakavandi and Z. Alaee, *Korean J. Chem. Eng.*, **33**, 1352 (2016).
20. Y. Ammi, L. Khaouane and S. Hanini, *Korean J. Chem. Eng.*, **32**, 2300 (2015).
21. Satya Eswari J. and N. Chandrakar, *Korean J. Chem. Eng.*, **33**, 1318 (2016).
22. S. Atashrouz and H. Mirshekar, *Bulg. Chem. Commun.*, **46**, 104 (2014).
23. S. Atashrouz, G. Pazuki and Y. Alimoradi, *Fluid Phase Equilib.*, **372**, 43 (2014).
24. M. Hashemkhani, R. Soleimani, H. Fazeli, M. Lee, A. Bahadori and M. Tavalaeian, *J. Mol. Liq.*, **211**, 534 (2015).
25. A. Hemmati-Sarapardeh, B. Aminshahidy, A. Pajouhandeh, S. H. Yousefi and S. A. Hosseini-Kaldozakh, *J. Taiwan Inst. Chem. Eng.*, **59**, 1 (2016).
26. C.-M. Vong, P.-K. Wong and Y.-P. Li, *Eng. Appl. Artif. Intell.*, **19**, 277 (2006).
27. A. Hemmati-Sarapardeh, R. Alipour-Yeganeh-Marand, A. Naseri, A. Safiabadi, F. Gharagheizi, P. Ilani-Kashkouli and A. H. Mohammadi, *Fluid Phase Equilib.*, **354**, 177 (2013).
28. A. Eslamimanesh, F. Gharagheizi, M. Illbeigi, A. H. Mohammadi, A. Fazlali and D. Richon, *Fluid Phase Equilib.*, **316**, 34 (2012).
29. M. Hosseinzadeh and A. Hemmati-Sarapardeh, *J. Mol. Liq.*, **200**, 340 (2014).
30. A. G. Ivakhnenko, *Sov. Autom. Control.*, **13**, 43 (1968).
31. A. G. Ivakhnenko, *IEEE Trans. Syst. Man. Cybern.*, 364 (1971).
32. S. Z. Reyhani, H. Ghanadzadeh, L. Puigjaner and F. Recances, *Ind. Eng. Chem. Res.*, **48**, 2129 (2009).
33. S. Atashrouz, G. Pazuki and S. S. Kakhki, *J. Mol. Liq.*, **202**, 95 (2015).

34. S. Atashrouz, E. Amini and G. Pazuki, *Ionics (Kiel)*, **21**, 1595 (2014).
35. V. N. Vapnik, Statistical Learning Theory, Wiley (1998).
36. J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor and J. Vandewalle, Least Squares Support Vector Machines, World Scientific (2002).
37. J. Suykens and J. Vandewalle, *Neural Process Lett.*, **9**, 293 (1999).
38. R. L. Haupt and S. E. Haupt, Practical genetic algorithms, Wiley (2004).
39. K. K. Tahboub, M. Barghash, M. Arafeh and O. Ghazal, *Math. Probl. Eng.*, **2016**, 1 (2016).
40. H. Karimi, F. Yousefi and M. R. Rahimi, *Heat Mass Transf.*, **47**, 1417 (2011).
41. S. Atashrouz, M. Mozaffarian and G. Pazuki, *Korean J. Chem. Eng.*, Accepted Manuscript (2016), DOI:10.1007/s11814-016-0169-4.
42. R. Ge, C. Hardacre, P. Nancarrow and D. W. Rooney, *J. Chem. Eng. Data*, **52**, 1819 (2007).
43. M. Tariq, M. G. Freire, B. Saramago, J. A. P. Coutinho, J. N. C. Lopes and L. P. N. Rebelo, *Chem. Soc. Rev.*, **41**, 829 (2012).
44. P. Navia, J. Troncoso and L. Romanı, *J. Chem. Eng. Data*, **52**, 1369 (2007).
45. H. Ning, M. Hou, Q. Mei, Y. Liu, D. Yang and B. Han, *Sci. China Chem.*, **55**, 1509 (2012), DOI:10.1007/s11426-012-4655-1.
46. I. Bahadur, N. Deenadayalu, Z. Tywabi, S. Sen and T. Hofman, *J. Chem. Thermodyn.*, **49**, 24 (2012).
47. A. B. Pereiro and A. Rodríguez, *J. Chem. Eng. Data*, **52**, 600 (2007).
48. S. Singh, I. Bahadur, G. G. Redhi, E. E. Ebenso and D. Ramjugernath, *J. Chem. Thermodyn.*, **89**, 104 (2015).
49. S. Singh, I. Bahadur, G. G. Redhi, E. E. Ebenso and D. Ramjugernath, *J. Mol. Liq.*, **199**, 518 (2014).
50. S. Wang, J. Jacquemin, P. Husson, C. Hardacre and M. F. C. Gomes, *J. Chem. Thermodyn.*, **41**, 1206 (2009).
51. S. Singh, I. Bahadur, G. G. Redhi, D. Ramjugernath and E. E. Ebenso, *J. Mol. Liq.*, **200**, 160 (2014).
52. A. B. Pereiro, E. Tojo, A. Rodríguez, J. Canosa and J. Tojo, *J. Chem. Thermodyn.*, **38**, 651 (2006).
53. E. Rilo, J. Pico, S. García-Garabal, L. M. Varela and O. Cabeza, *Fluid Phase Equilib.*, **285**, 83 (2009).
54. E. Gomez, B. Gonzalez, N. Calvar and A. Dominguez, *J. Chem. Thermodyn.*, **40**, 1208 (2008).
55. H. Rodríguez and J. Brennecke, *J. Chem. Eng. Data*, **51**, 2145 (2006).
56. Y. Tian, X. Wang and J. Wang, *Engineering*, **53**, 2056 (2008).
57. A. B. Pereiro and A. Rodriguez, *Phys. Chem. Liq.*, **46**, 172 (2008).
58. E. Gómez, B. González, Á. Domínguez, E. Tojo and J. Tojo, *J. Chem. Eng. Data*, **51**, 696 (2006).
59. A. Tekin, J. Safarov, A. Shahverdiyev and E. Hassel, *J. Chem. Eng. Data*, **136**, 177 (2007).
60. G. García-Miaja, J. Troncoso and L. Romaní, *J. Chem. Eng. Data*, **52**, 2261 (2007), DOI:10.1021/je7002836.
61. D. Waliszewski and H. Piekarski, *J. Chem. Thermodyn.*, **42**, 189 (2010).
62. D. Waliszewski, *J. Chem. Thermodyn.*, **40**, 203 (2008).
63. Y. H. Yu, A. N. Soriano and M. H. Li, *J. Taiwan Inst. Chem. Eng.*, **40**, 205 (2009).
64. P. Y. Lin, A. N. Soriano, A. R. Caparanga and M. H. Li, *Thermochim. Acta.*, **496**, 105 (2009).
65. Y. H. Yu, A. N. Soriano and M. H. Li, *Thermochim. Acta.*, **482**, 42 (2009).
66. W. Chen, L. Qiu, S. Liang, X. Zheng and D. Tang, *Thermochim. Acta.*, **560**, 1 (2013).
67. A. Wandschneider, J. K. Lehmann and A. Heintz, *J. Chem. Eng. Data*, **53**, 596 (2008).
68. M. Geppert-Rybczynska, J. K. Lehmann and A. Heintz, *J. Chem. Eng. Data*, **56**, 1443 (2011).
69. M. Geppert-Rybczyńska, J. K. Lehmann, J. Safarov and A. Heintz, *J. Chem. Thermodyn.*, **62**, 104 (2013).
70. E. Rilo, M. Dominguez-Perez, J. Vila, L. M. Varela and O. Cabeza, *J. Chem. Thermodyn.*, **49**, 165 (2012).
71. U. Domańska, A. Pobudkowska and M. Rogalski, *J. Colloid Interface Sci.*, **322**, 342 (2008).
72. J. Y. Wang, F. Y. Zhao, Y. M. Liu, X. L. Wang and Y. Q. Hu, *Fluid Phase Equilib.*, **305**, 114 (2011).
73. J. W. Russo and M. M. Hoffmann, *J. Chem. Eng. Data*, **56**, 3703 (2011).
74. J. S. Torrecilla, T. Rafione, J. García and F. Rodríguez, *J. Chem. Eng. Data*, **53**, 923 (2008).
75. J. Tong, M. Hong, Y. Chen, H. Wang, W. Guan and J.-Z. Yang, *J. Chem. Thermodyn.*, **54**, 352 (2012).
76. J. Y. Wang, X. J. Zhang, Y. Q. Hu, G. Di Qi and L. Y. Liang, *J. Chem. Thermodyn.*, **45**, 43 (2012).
77. H. F. D. Almeida, J. A. Lopes-da-Silva, M. G. Freire and J. A. P. Coutinho, *J. Chem. Thermodyn.*, **57**, 372 (2013).

Supporting Information

Implementation of soft computing approaches for prediction of physicochemical properties of ionic liquid mixtures

Saeid Atashrouz*, Hamed Mirshekar, Abdolhossein Hemmati-Sarapardeh***, Mostafa Keshavarz Moraveji****,†, and Bahram Nasernejad******

*Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Mahshahr Campus, Mahshahr, Iran
**Iran Polymer and Petrochemical Institute (IPPI), Tehran, Iran
***Department of Petroleum Engineering, Amirkabir University of Technology, Tehran, Iran
****Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez 424, P. O. Box 15875-4413, Tehran, Iran
(*Received 4 May 2016 • accepted 30 September 2016*)

Table S1. Optimized GMDH-PNN model for prediction of density

$N_1=-0.0719736+0.556177x_2+0.917127x_2x_4-1.06626x_2x_5-0.359876\ (x_2)^2-0.125207x_4+1.13837x_5$
$N_2=0.521624+0.00274322x_1x_5-5.08883e\text{-}06\ (x_1)^2-0.158605x_2x_3-1.0043x_2x_5+1.15338x_2N_1+0.127511x_3x_5$
$N_3=0.22599-0.785497x_2x_5+0.779865x_2N_2-0.225048x_3x_4+0.216699x_3N_2+0.198607x_4x_5+0.585891x_5$
Density $(g\cdot cm^{-3})=-0.264718+0.640599x_5-0.267305\ (x_5)^2-1.01067N_1+0.460895N_1^2-0.368188N_2N_3+1.80215N_3$

Note: x_1=temperature (kelvin), x_2=mole fraction of IL, x_3=mole fraction of second compound, x_4=density of pure IL ($g\cdot cm^{-3}$), x_5=density of second compound ($g\cdot cm^{-3}$)

Table S2. Optimized GMDH-PNN model for prediction of heat capacity

$N_1=213.594-0.997025x_3x_4+0.994389x_3x_5+0.00446013\ (x_4)^2-0.685411x_5+0.000733119\ (x_5)^2$
$N_2=237.966+0.0496782x_2x_1-11.8592\ (x_2)^2-1.56183x_1+0.00250384\ (x_1)^2+1.00414N_1$
$N_3=-0.6521+0.0306576x_4N_1-0.030601x_4N_2-0.0113466x_1N_1+x_1N_20.0112716+1.01965N_1$
Heat capacity $(J\cdot mol^{-1}\cdot K^{-1})=-2.12811-2.63109\ (x_3)^2+0.0923764x_4N_1-0.0922488x_4N_3-7.10469N_1+8.10501N_3$

Note: x_1=temperature (kelvin), x_2=mole fraction of IL, x_3=mole fraction of second compound, x_4=heat capacity of pure IL ($J\cdot mol^{-1}\cdot K^{-1}$), x_5=heat capacity of second compound ($J\cdot mol^{-1}\cdot K^{-1}$)

Table S3. Optimized GMDH-PNN model for prediction of thermal conductivity

$N_1=0.0863986+0.867651x_2x_4-1.22141x_2x_5+0.114389\ (x_2)^2+2.73967\ (x_5)^2$
Thermal conductivity $(W\cdot m^{-1}\cdot K^{-1})=-0.0525781+0.00561402x_1N_1-1.61322x_2x_5+1.72411x_2N_1+1.53019x_5N_1-3.48679N_1{}^2$

Note: x_1=temperature (kelvin), x_2=mole fraction of IL, x_3=mole fraction of second compound, x_4=thermal conductivity of pure IL ($W\cdot m^{-1}\cdot K^{-1}$), x_5=thermal conductivity of second compound ($W\cdot m^{-1}\cdot K^{-1}$)

Table S4. Optimized GMDH-PNN model for prediction of surface tension of organic based mixtures

$N_1=-11.0344-0.822476x_3x_4+0.857515x_3x_5+1.92507x_4-0.017799(x_4)^2$
$N_2=42.0285-0.997937x_4+0.0299199x_4x_5-0.754072x_5+0.0145256N_1^2$
$N_3=8.54067-14.2635x_2+0.0567047x_2x_5+0.358895x_2N_2+0.75127N_2$
Surface tension$=-7.7141+0.309498x_4-0.0537704x_4N_2+0.0448101x_4N_3+1.22457N_2$

Note: x_1=temperature (kelvin), x_2=mole fraction of IL, x_3=mole fraction of second compound, x_4=surface tension of pure IL ($mN\cdot m^{-1}$), x_5=surface tension of second compound ($mN\cdot m^{-1}$)

Table S5. Optimized GMDH-PNN model for prediction of surface tension of water based mixtures

$$N_1 = 79992 - 0.709299x_1x_4 - 0.167555\ (x_1)^2 - 692.774x_4 + 0.0196615\ (x_5)^2$$

$$N_2 = -30.0526 + 1.85817x_3x_5 - 1.85783x_3N_1 - 0.0183048x_5N_1 + 2.51251N_1$$

$$N_3 = -2.73914 + 43.669x_1x_2 - 31475.8x_2 + 256.377x_2x_4 + 1.06082N_2$$

$$\text{Surface tension} = 11.1383 + 0.0863794x_4N_2 - 0.0783936x_4N_3 - 0.144503N_2N_3 + 0.148288N_3^2$$

Note: x_1=temperature (kelvin), x_2=mole fraction of IL, x_3=mole fraction of second compound, x_4=surface tension of pure IL ($mN \cdot m^{-1}$), x_5= surface tension of second compound ($mN \cdot m^{-1}$)