



OPEN

Solubility of gaseous hydrocarbons in ionic liquids using equations of state and machine learning approaches

Reza Nakhaei-Kohani¹, Saeid Atashrouz²✉, Fahimeh Hadavimoghaddam^{3,4}, Ali Bostani⁵, Abdolhossein Hemmati-Sarapardeh⁶✉ & Ahmad Mohaddespour⁷✉

Ionic liquids (ILs) have emerged as suitable options for gas storage applications over the past decade. Consequently, accurate prediction of gas solubility in ILs is crucial for their application in the industry. In this study, four intelligent techniques including Extreme Learning Machine (ELM), Deep Belief Network (DBN), Multivariate Adaptive Regression Splines (MARS), and Boosting-Support Vector Regression (Boost-SVR) have been proposed to estimate the solubility of some gaseous hydrocarbons in ILs based on two distinct methods. In the first method, the thermodynamic properties of hydrocarbons and ILs were used as input parameters, while in the second method, the chemical structure of ILs and hydrocarbons along with temperature and pressure were used. The results show that in the first method, the DBN model with root mean square error (RMSE) and coefficient of determination (R^2) values of 0.0054 and 0.9961, respectively, and in the second method, the DBN model with RMSE and R^2 values of 0.0065 and 0.9943, respectively, have the most accurate predictions. To evaluate the performance of intelligent models, the obtained results were compared with previous studies and equations of the state including Peng–Robinson (PR), Soave–Redlich–Kwong (SRK), Redlich–Kwong (RK), and Zudkevitch–Joffe (ZJ). Findings show that intelligent models have high accuracy compared to equations of state. Finally, the investigation of the effect of different factors such as alkyl chain length, type of anion and cation, pressure, temperature, and type of hydrocarbon on the solubility of gaseous hydrocarbons in ILs shows that pressure and temperature have a direct and inverse effect on increasing the solubility of gaseous hydrocarbons in ILs, respectively. Also, the evaluation of the effect of hydrocarbon type shows that increasing the molecular weight of hydrocarbons increases the solubility of gaseous hydrocarbons in ILs.

Abbreviations

ELM	Extreme learning machine
DBN	Deep belief network
B-SVR	Boosting support vector regression
MARS	Multivariate adaptive regression splines
RBM	Restricted Boltzmann machine
PR	Peng–Robinson
SRK	Soave–Redlich–Kwong
RK	Redlich–Kwong
ZJ	Zudkevitch–Joffe

¹Department of Chemical and Petroleum Engineering, Shiraz University, Shiraz, Iran. ²Department of Chemical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran. ³Key Laboratory of Continental Shale Hydrocarbon Accumulation and Efficient Development (Northeast Petroleum University), Ministry of Education, Northeast Petroleum University, Daqing 163318, Heilongjiang, China. ⁴Institute of Unconventional Oil and Gas, Northeast Petroleum University, Daqing 163318, China. ⁵College of Engineering and Applied Sciences, American University of Kuwait, AUK, P.O. Box 3323, Salmiya, Kuwait. ⁶Department of Petroleum Engineering, Shahid Bahonar University of Kerman, Kerman, Iran. ⁷Department of Chemical Engineering, McGill University, Montreal, QC H3A 0C5, Canada. ✉email: s.atashrouz@gmail.com; hemmati@uk.ac.ir; ahmad.mohaddespour@mail.mcgill.ca

EOS	Equation of state
RMSE	Root mean squared error
R ²	Coefficient of determination

Ionic liquids (ILs), unlike most salts, are substances that are liquid at temperatures lower than 373.15 K. As a result of various combinations of cations and anions, numerous ILs with tunable properties can be obtained^{1,2}. ILs have become commonly used in different industries due to their various appealing properties, such as thermal stability, non-flammability, and various electrical and ionic conductivity. Therefore, many ILs have been synthesized for various purposes in recent decades^{3,4}. The implementation of effective solvents is critical in the battle to conserve the environment and eliminate dangerous atmospheric concentrations^{5,6}. ILs are one of the most prominent eco-friendly solvents. ILs might be regarded as attractive compounds with various potential industrial uses because of these exceptional properties^{5–7}. Recent works on ILs have concentrated on structure–property correlations to intelligently develop ILs for special purposes. The structure–property interactions related to gas solubility in IL are of considerable importance. The application of ILs in separation processes highlights the relevance of gas solubility in ILs. Acknowledging the dissolution process of substances in ILs and how they might investigate the key structural properties of the solvent seems to be a more basic objective^{8,9}. For many purposes, employing ILs to absorb the major components of natural gas has been a popular study subject. Primarily, such information is required to investigate the natural gas sweetening operation, which involves the separation of acid gases from methane. This technique is important not only for ecological reasons, but also for preventing subsequent process issues including equipment corrosion and pipeline clogging^{5,10}. Furthermore, such information is useful for a variety of operations, including supercritical fluid extraction and homogeneous interactions in ILs, as well as purification of steam converting or gas-shift processes^{3,5}. Since ILs are structurally heterogeneous on a very small scale, the solubility of substances may differ from that of molecular liquids. The solubility of pure gaseous hydrocarbons is also a promising topic for investigating the performance of nonpolar component solvation in the ILs. While some experimental researches of gaseous hydrocarbons solubility in ILs have been conducted, the majority of these researches have relied on gas solubility in ILs based on thermodynamic properties⁸.

Therefore, the study of the solubility of various gases, especially gaseous hydrocarbons in ILs is of special importance. There are various methods such as laboratory tests and the employment of equations of state to measure the solubility of gaseous hydrocarbons in ILs. Many studies have been performed on the solubility of methane^{5,11}, ethane^{12,13}, normal butane and isobutene⁸, and other gaseous hydrocarbons in various ILs. Ahsan Jalal¹⁴ calculated the solubilities of C4-hydrocarbons in various imidazolium-type ILs using the Conductor-like Screening Model for Realistic Solvents (COSMO-RS) computations. The COSMO-RS data for each hydrocarbon was then examined utilizing semi-empirically determined molecular descriptors of ILs and machine learning technologies such as association rule mining and decision tree categorization. The most essential characteristics for defining the propensity of ILs towards C4-hydrocarbons are the polarizabilities of both cation and anion, as well as the anion's CPK (space-filling model) area. Zhen Song¹⁵ used COSMO-RS to forecast the solubilities of 220 imidazolium-based ILs in seven fuel hydrocarbon samples with varying IL and hydrocarbon properties. The nitrogen analyzer was used to measure the solubilities of five typical ILs in the examined hydrocarbons, revealing COSMO-RS's adequate capacity to forecast IL-in-hydrocarbon solubilities. ALSaleem¹⁶ investigated the solubility of three halogenated hydrocarbons in 12 hydrophobic ILs at temperatures ranging from 25 to 45 °C. Piperidinium, pyrrolidinium, and ammonium-based cations coupled with [TF₂N] anion were studied for their chemical structure and alkyl chain length impact. Carbon tetrachloride and bromoform are somewhat miscible in all of the ILs tested, although chloroform is fully miscible. The solubility of ammonium-based ionic liquids rises as the cation molecular weight and alkyl chain length increase. Hamed et al.⁵ calculated the solubility of methane in ILs using two equations of state including modified Sanchez and Lacombe (SL) and cubic-plus-association (CPA). Comparison of the results showed that the mentioned methods have good accuracy in predicting the solubility of methane in ILs. Owing to the very non-ideal reactions, thermodynamic analysis of gas solubility in ILs is a difficult undertaking. Typical EOSs, on the other hand, overlook the influence of several essential non-ideal molecular interactions, like hydrogen bonds and polar interactions, and so are unreliable in predicting gas solubility in ILs^{5,17}. This demands the evaluation of more powerful EOSs, which are capable of adequately modeling the system across a broad temperature and pressure range and have reasonable conformance with the real performance of the system. Recently, machine learning-based approaches have widely been received attention for predicting properties of ILs^{18–24}. The excellent results of previous studies further demonstrate the reliability and applicability of these methods. Other advantages of these methods are their high accuracy in predicting the solubility of gases in ILs and reducing time and cost. Ferreira et al.²⁵ conducted a study on binary systems of ionic liquids and liquid hydrocarbons and their modeling by COSMO-RS. The results showed that COSMO-RS can be a useful and practical predictive tool. The employment of intelligent methods to predict the solubility of various gases such as N₂O²⁶, SO₂^{27,28}, H₂S^{24,29}, and CO₂³⁰ in ILs has received much attention in the last decade, and the results confirm the reliability of these methods. Using the thermodynamic properties of ILs and 728 experimental data, Baghban et al.³¹ developed a multi-layer perceptron (MLP) model and an adaptive neuro-fuzzy interference system (ANFIS) to predict the solubility of CO₂ in ILs and then compared the results with the equations of state. The results showed that the MLP model had the best performance with mean square error (MSE) and coefficient of determination (R²) values of 0.000133 and 0.9972, respectively. Ahmadi et al.³² did the same work by developing artificial neural networks optimized with backpropagation (BP) and particle swarm optimization (PSO) algorithms to predict the solubility of H₂S in ILs. The mean square error (MSE) and R² values for PSO-ANN were 0.00025 and 0.99218, respectively, which indicated the better performance of this model than BP-ANN. Song et al.³³ predicted the solubility of CO₂ in various ILs by using a very large database containing 10,116 experimental data points. In this study, artificial neural networks and support vector machine

(SVM) were used to develop the group contribution (GC) method. The results showed that the ANN-GC model had the best performance with MSE and R^2 values of 0.0202 and 0.9836, respectively. In 2021, Moosanezhad-Kermani et al.³⁴ developed a group method of data handling method for prediction CO_2 solubility in ILs. In 2022, Mohammadi et al.³⁵ predicted SO_2 solubility in ILs using four intelligent models and equations of state.

In this study, using an extensive data bank including 2145 experimental data points, four intelligent models including Extreme Learning Machine (ELM), Deep Belief Network (DBN), Multivariate Adaptive Regression Splines (MARS), and Boosting-Support Vector Regression (B-SVR) are developed to predict the solubility of gaseous hydrocarbons (CH_4 , C_2H_6 , C_3H_8 , C_4H_{10} , C_2H_4 , C_3H_6 , C_4H_8 , and C_6H_6) in ILs in a wide range of temperature and pressure. In this regard, intelligent models are trained based on two strategies, including Model (I): based on the thermodynamic properties of ILs and hydrocarbons, and Model (II): based on the chemical structure of ILs and hydrocarbons. In the first method, the critical properties of ILs and hydrocarbons and their molecular weights are considered as input parameters to the model. In the second method, the effect of the substructures that make up ILs and gaseous hydrocarbons are considered. In the development of existing models, the data is divided into an 80/20 ratio for the training and testing stages. Results are compared with previous studies and equations of state to evaluate the precision of the developed model. Furthermore, the solubility trend study is also performed to investigate the effect of different parameters on the solubility of gaseous hydrocarbons in ILs.

Data collection. In this study, 2145 experimental data for the solubility of gaseous hydrocarbons (CH_4 , C_2H_6 , C_3H_8 , C_4H_{10} , C_2H_4 , C_3H_6 , C_4H_8 , and C_6H_6) in ionic liquids over a wide range of temperature (280–453.15 K) and pressure (0–201.64 bar) have been collected from the literature^{3,4,12,13,36–54}. The details of the provided database are summarized in Table 1. In the method based on thermodynamic properties, temperature, pressure, molecular weight, boiling point, and critical properties of ILs and hydrocarbons are considered as input parameters, while in the method based on chemical structure, in addition to the substructures of ILs and hydrocarbons, temperature and pressure are considered as input parameters. The statistical description of the input parameters and a number of substructures used in this study are reported in Table 2 and Table 3, respectively. The data are then divided into an 80/20 ratio in the training and testing stages, respectively. Figure 1 illustrates the chemical structure of anions and cations used in this study.

Modeling

Multivariate adaptive regression spline (MARS). The multivariate adaptive regression spline (MARS), which is comprised of a count of different basis functions, will be used for regressive assessment and the identification of estimation techniques. The MARS model can calculate the enhanced structure of the multiplication of spline basis functions (BFs). The MARS system may choose the count of BFs and the properties related to each one, like the multiplication degree and node positions, autonomously. We might mention recurrent regression as one of the various clustering techniques. To properly describe higher-order responses, the MARS model employs the backward stepwise approach^{55,56}. The MARS concept has several strengths. These benefits provide several items that will be detailed further below. This model's practical property is that the count of parameters in this mathematical equation is reduced, which speeds up the calculation of the body while reducing simulation precision. Conserving computing resources is extremely desired for researchers. This model's clever method of dividing factors into two groups of crucial factors and supplementary factors, which maximize detection accuracy, is another noteworthy aspect. The mathematical look of this concept is one of its most appealing qualities⁵⁷. The dependent and independent components are identified by the mathematical expression form⁵⁸. There is no supposition for linking inputs and output among the premises utilized in the MARS system. The purpose of this analysis is to look into the relationship between the output y and inputs $\{x_1, \dots, x_n\}$ based on previous revelations. The following is a description of the data generation system⁵⁹:

$$y = f(x_1, \dots, x_n) + e \quad (1)$$

e refers to the relative error that matches across the intended region (x_1, \dots, x_n) in which the MARS model is generated. A progressive linear regression technique is a frequent model-building approach that MARS employs. MARS employs an automated way to replace major inputs. A succession of comparable splines and their derivatives is created to provide these possibilities. The splines are coupled and form polynomial curves to construct an ideal model. BFs are the name for these polynomials. They can simulate both linear and nonlinear processes. BFs are often split into couples, each piecewise linear or cubic in its own right. They have nodes at a certain place, each of which is related to the inputs. Below equation gives the shape of distributed linear BFs as⁵⁹:

$$(x - t)_+ = \max(0, x - t) = \begin{cases} x - t & \text{if } x > t \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad (t - x)_+ = \max(0, t - x) = \begin{cases} t - x & \text{if } x < t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The positive component is denoted by “+”, while the node is denoted by “t”. The MARS model is calculated using the following formula in its standard form⁵⁹:

$$f(X) = \beta_0 + \sum_{m=1}^M \beta_m B_m(X) \quad (3)$$

In the above equation, X is the input vector, and B_m , β_m , and β_0 are m th BF, coefficient of m th BF, and the constant of the equation, respectively.

For the MARS approach, there are two stages to follow⁵⁹:

IL full name	IL abbreviation	Hydrocarbon	T (K)	P (bar)	S (mole fraction)	N
1-Butyl-3-methylimidazolium tetrafluoroborate	[BMIM][BF ₄]	CH ₄	283.05–343.09	0.465–0.976	0.00045–0.00126	13
1-Ethyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[EMIM][eFAP]	CH ₄	293.3–363.42	20.76–86.92	0.052–0.155	31
1-Ethyl-3-methylimidazolium ethylsulfate	[EMIM][EtSO ₄]	CH ₄	292.31–293.63	1.98–101.5	0.0013–0.0405	8
1-Butyl-3-methylimidazolium methylsulfate	[BMIM][MeSO ₄]	CH ₄	293.15–413.2	13.63–88.53	0.0091–0.046	24
1-Butyl-3-methylimidazolium hexafluorophosphate	[BMIM][PF ₆]	CH ₄	283.15–343.12	0.04–13.99	0–0.1138	99
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	CH ₄	300.13–453.15	15.01–161.05	0.0298–0.2245	124
1-Hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[HMIM][Tf ₂ N]	CH ₄	293.3–413.25	0.0158–93	0.0007–0.186	125
1-Hexyl-3-methylpyridinium bis(trifluoromethylsulfonyl) imide	[HMPY][Tf ₂ N]	CH ₄	298.15–333.15	0.0158–10	0.0007–0.0277	50
1,1,3,3-tetramethylguanidine lactate	[TMG][L]	CH ₄	308–328	10.4–103.4	0.0009–0.0431	30
1-Hexyl-3-methylimidazolium tricyanomethanide	[HMIM][TCM]	CH ₄	293.26–363.37	18–103.6	0.025–0.1	31
N-Methyl-(2-hydroxyethyl)amine propionate	[m2HEA][Pr]	CH ₄	313–363.1	25–153.9	0.0153–0.0784	44
Bis(2-hydroxyethyl)amine propionate	[BHEA][Pr]	CH ₄	333.1–363.1	48.2–114.6	0.0141–0.0537	16
(2-hydroxyethyl)amine propionate	[2HEA][Pr]	CH ₄	333.1–363.1	48.2–114.6	0.0141–0.0537	16
Bis(2-hydroxyethyl) ammonium butanoate	[BHEA][Bu]	CH ₄	313–353	24.26–201.64	0.018–0.073	20
1-Hexyl-3-methylimidazolium nitrate	[HMIM][NO ₃]	CH ₄	293.15–343.15	8.74–30.55	0.0204–0.0993	30
1-Ethyl-3-methylimidazolium diethylphosphate	[EMIM][dep]	CH ₄	303.17–363.29	16.85–94.41	0.02–0.076	35
Trihexyltetradecylphosphonium bis(2,4,4-trimethylpentyl)phosphinate	[thtdp][phos]	CH ₄	302–363.27	10.15–120.49	0.107–0.496	42
Trihexyltetradecylphosphonium dicyanamide	[thtdp][dca]	CH ₄	302.13–363.48	14.28–115.69	0.079–0.343	49
1-Allyl-3-methylimidazolium dicyanamide	[amim][dca]	CH ₄	303.19–363.68	33.51–95.9	0.015–0.034	28
1-Butyl-1-methylpyrrolidinium dicyanamide	[bmpyrr][dca]	CH ₄	303.22–363.66	25.59–68.98	0.019–0.041	28
1,2,3-Tris(diethylamino)cyclopropenylum dicyanamide	[cprop][dca]	CH ₄	303.23–363.78	18.93–70.55	0.029–0.086	28
1,2,3-Tris(diethylamino)cyclopropenylum bis(trifluoromethylsulfonyl) imide	[cprop][Tf ₂ N]	CH ₄	302.82–363.42	18.8–78.85	0.066–0.19	35
1-Butyl-1-methylpiperidinium bis(trifluoromethylsulfonyl) imide	[bmpip][Tf ₂ N]	CH ₄	303.03–363.66	14.11–73.83	0.033–0.12	35
Triethylsulfonium bis(trifluoromethylsulfonyl)imide	[tes][Tf ₂ N]	CH ₄	303.1–363.46	12.46–82.3	0.024–0.111	35
Methyltriocylammonium bis(trifluoromethylsulfonyl) imide	[toa][Tf ₂ N]	CH ₄	302.96–363.72	11.65–75.43	0.076–0.29	35
1-Hexyl-3-methylpyridinium bis(trifluoromethylsulfonyl) imide	[HMPY][Tf ₂ N]	CH ₄	298.15–333.15	0.0158–10	0.0007–0.0277	50
1-Butyl-3-methylimidazolium tetrafluoroborate	[BMIM][BF ₄]	C ₂ H ₆	283.02–343.22	0.423–0.936	0.00198–0.00388	12
1-Butyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[BMIM][eFAP]	C ₂ H ₆	303.26–343.33	0.9437–1.083	0.0102–0.0173	15
1-Butyl-3-methylimidazolium hexafluorophosphate	[BMIM][PF ₆]	C ₂ H ₆	283.15–343.12	0.0002–13	0–0.043	172
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	C ₂ H ₆	283.15–323.15	0.000425–13	0–0.126	70
1-Butyl-1-methylpyrrolidinium tris(pentafluoroethyl)trifluorophosphate	[BMPYR][eFAP]	C ₂ H ₆	303.16–344.69	0.3887–0.92	0.00573–0.013	12
1-Ethyl-3-methylimidazolium ethylsulfate	[EMIM][EtSO ₄]	C ₂ H ₆	322.76–349.98	2.1–40.06	0.0016–0.0369	13
1-Ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[EMIM][Tf ₂ N]	C ₂ H ₆	304.14–344.7	0.4422–0.4977	0.0038–0.0066	9
1-Hexyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[HMIM][eFAP]	C ₂ H ₆	303.14–343.14	0.5204–0.6656	0.0091–0.0167	13
1-Hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[HMIM][Tf ₂ N]	C ₂ H ₆	283.32–368.4	0.4404–130.7	0.0085–0.4016	118
1-Hexyl-3-methylpyridinium bis(trifluoromethylsulfonyl)imide	[HMPY][Tf ₂ N]	C ₂ H ₆	298.15–333.15	0.0138–10	0.00076–0.132	54
Hydroxyethyl-propyl-dimethylammonium bis[(trifluoromethyl)sulfonyl] imide	[N1,1,3,2-OH][Tf ₂ N]	C ₂ H ₆	304.15–344.74	0.4363–0.4905	0.0036–0.0062	9
Trihexyl(tetradecyl) phosphonium tris(pentafluoroethyl) trifluorophosphate	[P6,6,6,14][eFAP]	C ₂ H ₆	303.16–343.29	0.6838–0.8019	0.0253–0.0414	15
N-Pentyl-N-Methylpyrrolidiniumbis(trifluoromethylsulfonyl)imide	[pmpyrr][Tf ₂ N]	C ₂ H ₆	298.15–333.15	0.1–13	0.0024–0.1404	30
Diethylmethyl(2-methoxyethyl)ammonium bis(trifluoromethylsulfonyl) imide	[deme][Tf ₂ N]	C ₂ H ₆	298.15–333.15	0.1–13	0.0023–0.1054	30
1,2,3-Tris(diethylamino)cyclopropenylum bis(trifluoromethanesulfonyl) imide	[TDC][Tf ₂ N]	C ₂ H ₆	298.15–333.15	0.1–13	0.0033–0.2029	30
1-Hexyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[HMIM][eFAP]	C ₂ H ₆	298.15–333.15	0.1–13	0.0034–0.195	30
1-Butyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[BMIM][eFAP]	C ₂ H ₆	298.15–333.15	0.1–13	0.002–0.1499	30
1-Ethyl-3-methylimidazolium tris(pentafluoroethyl)trifluorophosphate	[EMIM][eFAP]	C ₂ H ₆	298.15–333.15	0.1–13	0.0019–0.1318	30
1-Hexyl-3-methylimidazolium bis(trifluoromethylsulfonyl) imide	[HMIM][Tf ₂ N]	C ₃ H ₈	279.98–339.97	0.999–12.182	0.0506–0.2946	33
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	C ₄ H ₁₀	280–340	0.21–3.511	0.0284–0.1663	16
1-Butyl-3-methylimidazolium hexafluorophosphate	[BMIM][PF ₆]	C ₂ H ₄	283.15–323.15	0.01–13	0.00006–0.099	149
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	C ₂ H ₄	283.15–323.15	0.0007–13	0.0000005–0.176	57
1-Hexyl-3-methylpyridinium bis(trifluoromethylsulfonyl) imide	[HMPY][Tf ₂ N]	C ₂ H ₄	298.15–333.15	0.0134–10	0.00057–0.146	50
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	C ₃ H ₆	279.98–339.97	0.883–10.622	0.016–0.1818	16
1-Butyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide	[BMIM][Tf ₂ N]	C ₄ H ₈	280–340	0.334–3.041	0.0252–0.1409	16
1-Butyl-3-methylimidazolium tetrafluoroborate	[BMIM][BF ₄]	C ₆ H ₆	283.15–323.15	0.00001–0.202	0–0.43	55

Table 1. Description of systems of ionic liquids and hydrocarbons used in this study.

Property	Unit	Statistical parameters					
		Mean	Median	Kurtosis	Skewness	Min	Max
T	K	321.26	318	3.58	1.44	279.98	453.15
P	Bar	24.31	9.96	3.32	1.81	1E-05	201.64
Mw (Gas)	gr/mole	24.8	28.05	7.97	2.44	16.04	78.11
Tb (Gas)	K	153.3	169.5	4.05	1.64	111.63	353.3
Pc (Gas)	bar	47.45	46	1.99	-0.5	38	51.2
Tc (Gas)	K	255.7	283	3.83	1.57	190.55	563
ω (Gas)	-	0.06	0.086	0.065	0.75	0.011	0.212
Tc (IL)	K	1056.6	1189.9	-0.56	0.188	643.1	1878.8
Pc (IL)	bar	21	21.59	0.2	-0.2	5.13	40.46
Tb (IL)	K	772.11	839.7	0.4	0.57	495.2	1374.9
ω (IL)	-	0.59	0.585	0.111	0.072	-0.2	1.43
Mw (IL)	gr/mole	392.9	419.4	0.887	0.623	135.16	928.88
Solubility	Mole fraction	0.069	0.0367	5.2	2.19	5E-08	0.496

Table 2. Statistical description of databank used in the current study.

Substructures	
-CH ₃	-COO-
-CH ₂ -	=O
>CH-	-NH ₂
>C<	-NH ₃
=CH ₂	>N-
=CH-	=N-
=C<	-F
=C=	-P
≡C-	-B
-OH	-S-
-O-	-SO ₂

Table 3. Different substructures of ionic liquids and hydrocarbons.

Stage 1: In the targeted area, a collection of BFs is generated. This process begins with a structure that just has the intercept term and adds the phrases one at a time. BFs should be used for the phrases. The forward step can be stopped in two ways: the first is when the maximum number of BFs is reached, and the second is when the regression coefficient of modifications is less than the set value. Several BFs are consecutively utilized, resulting in an overfitting system.

Stage 2: Overfitting is undesired and must be avoided at all costs. Because the system uses a monitoring phrase, it must be condensed by deleting at least one more relevant BF at a time. Then, from among the best systems of any size, the system with the least general cross-validation (GCV) is chosen to complete the system. This procedure concludes with the creation of a final model. GCV is used in the MARS system to eliminate BFs that aren't needed. A GCV expression is as follows^{60,61}:

$$GCV = \frac{MSE_{train}}{\left(1 - \frac{enp}{N}\right)^2} \quad (4)$$

MSE_{train} represents the mean square error when utilizing training datasets, N represents the size of observations, and enp indicates the number of factors in the practical circumstance. Figure 2 illustrates the schematic of the MARS model.

Extreme learning machine (ELM). Huang G.B. invented the extreme learning machine (ELM) to minimize time-consuming repeated training and improve segmentation accuracy^{62,63}. The ELM architecture consists of an input layer, a concealed layer, and an output layer. It is a single-hidden-layer feedforward neural network (SLFN). Unlike conventional learning methods (like the Backpropagation), which arbitrarily established all system training variables and quickly build local best solutions, the ELM just adjusts the neurons in the invisible layer, randomizes the weights between the visible and invisible layers, and also the bias of the hidden neurons in the method implementation stage, determines the hidden layer output matrix, and subsequently finds the weights. The ELM provides the prospect of quick learning speed because of its straightforward network archi-

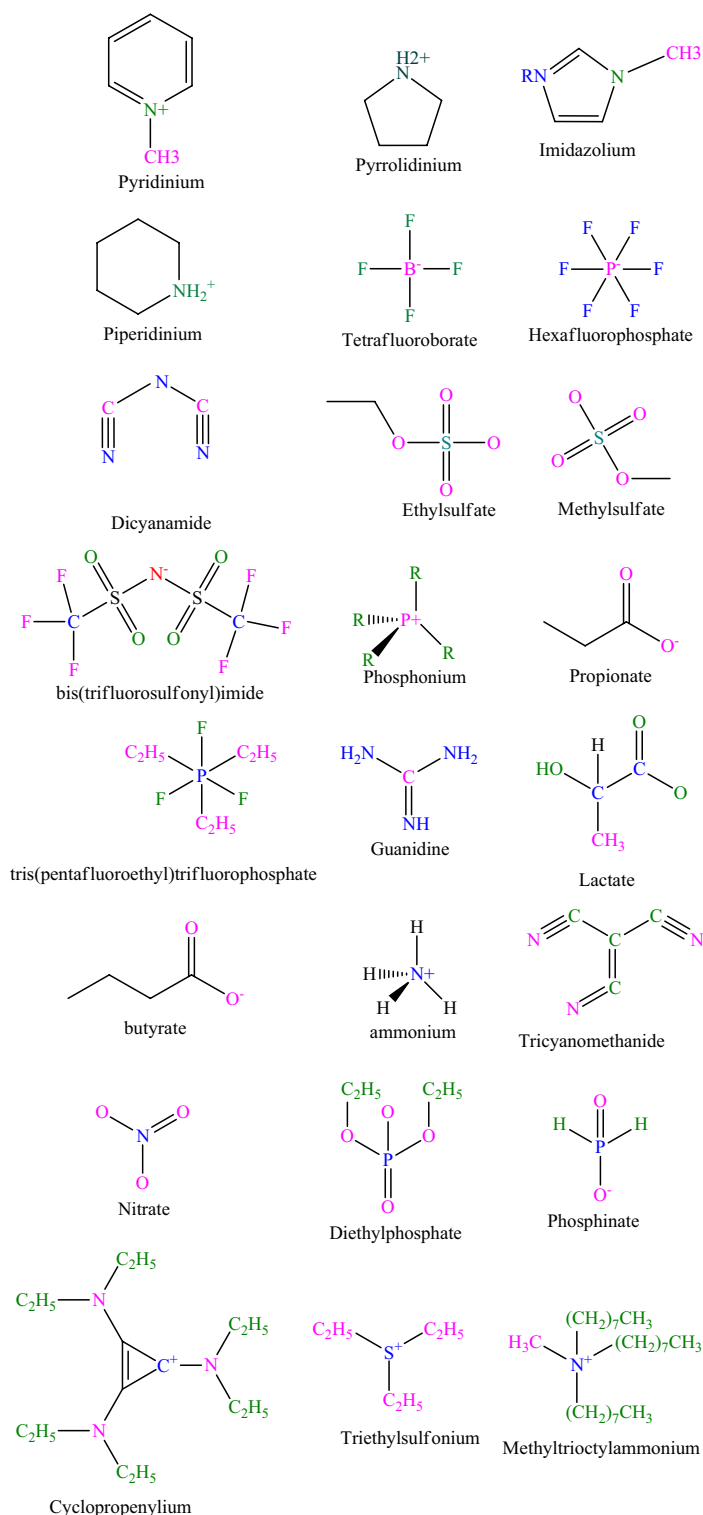


Figure 1. Chemical structure of some of anions and cations studied in this work.

texture and succinct variable calculation operations. Figure 3 depicts the architectural framework of ELM. Figure 3 shows the construction of an extreme learning machine system, which has n input layer neurons, l hidden layer neurons, and m output layer neurons. Given the training set $\{X, Y\} = \{x_i, y_i\}$ ($i = 1, \dots, Q$), with an input parameter $X = [x_{i1} x_{i2} \dots x_{iQ}]$ and an output matrix $Y = [y_{i1} y_{i2} \dots y_{iQ}]$ composed of the training set, the matrix X and the matrix Y may be represented as follows⁶⁴:

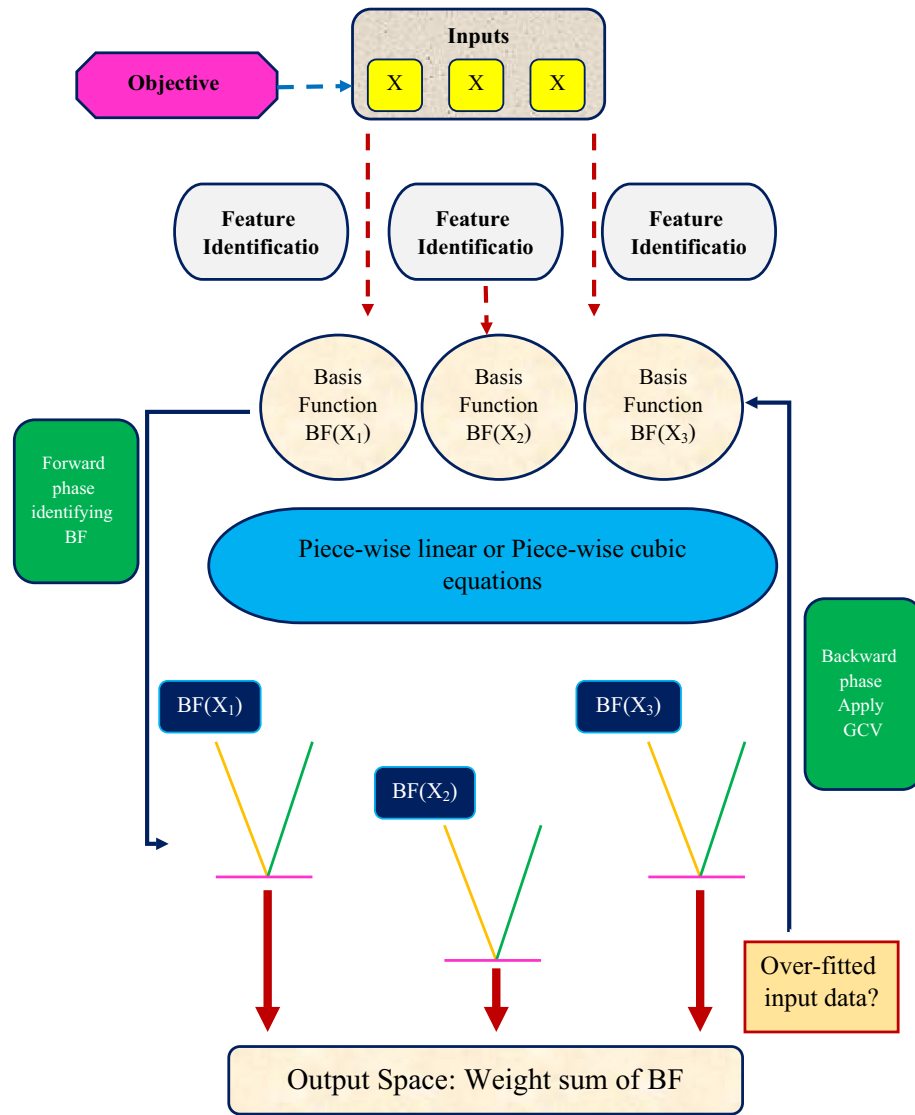


Figure 2. Schematic of the MARS model.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1Q} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nQ} \end{bmatrix}, Y = \begin{bmatrix} y_{11} & \cdots & y_{1Q} \\ \vdots & \ddots & \vdots \\ y_{m1} & \cdots & y_{mQ} \end{bmatrix} \tag{5}$$

here, n and m are the dimensions of the input and output matrices, respectively.

The ELM then adjusts the weights between the input and hidden layers at random⁶⁴:

$$w = \begin{bmatrix} w_{11} & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{l1} & \cdots & w_{ln} \end{bmatrix} \tag{6}$$

The weights between the jth input layer neuron and the ith hidden layer neuron are represented by w_{ij} . Third, the ELM makes the following assumptions about the weights between the hidden and output layers⁶⁴:

$$\beta = \begin{bmatrix} \beta_{11} & \cdots & \beta_{1m} \\ \vdots & \ddots & \vdots \\ \beta_{l1} & \cdots & \beta_{lm} \end{bmatrix} \tag{7}$$

The weights between the jth hidden layer neuron and the kth output layer neuron are represented by β_{jk} . Fourth, the ELM sets the buried layer neurons' bias at random⁶⁴:

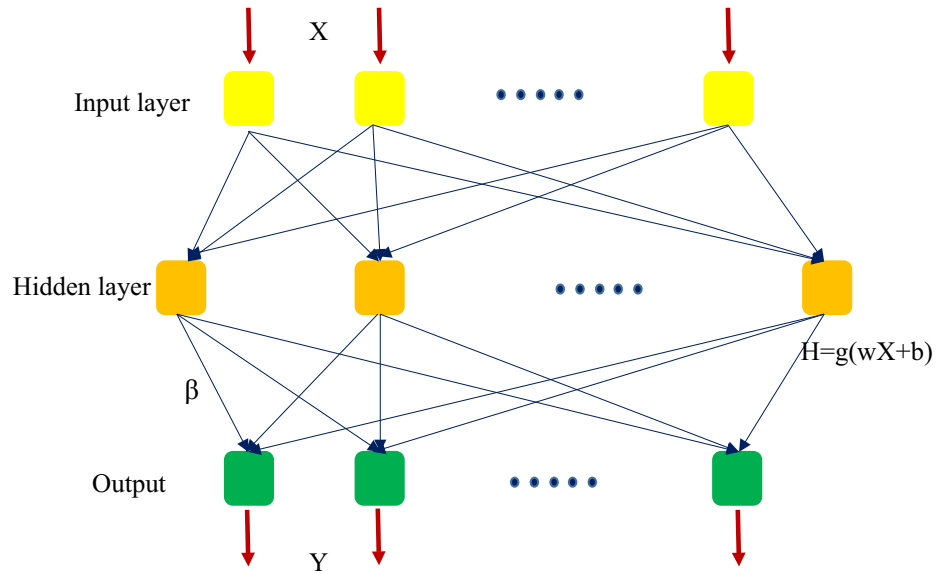


Figure 3. Schematic of the structure of ELM.

$$B = [b_1 b_2 \dots b_n]^T \tag{8}$$

Fifth, the ELM selects the $g(x)$ system activation function.

The output matrix T may be represented as follows according to Fig. 3:

$$T = [t_1, t_2, \dots, t_Q]_{m \times Q} \tag{9}$$

The output matrix T 's column vectors are as follows⁶⁴:

$$t_j = \begin{bmatrix} t_{1j} \\ \vdots \\ t_{mj} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^l \beta_{i1} g(w_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^l \beta_{im} g(w_i x_j + b_i) \end{bmatrix}, \quad j = 1, \dots, Q \tag{10}$$

Sixth, have a look at Eqs. (9) and (10) to see what we can come up with:

$$H\beta = T' \tag{11}$$

here, T' is the transpose of T and H is the concealed layer's output. We utilize the least square approach to determine the weight matrix values of β ⁶² to find a particular solution with the least amount of error.

$$\beta = H^+ T' \tag{12}$$

We apply a regularization element to the β ⁶⁵ to increase the network's generalization capabilities and make the findings more robust. When the count of concealed layer neurons is smaller than the number of training sets, β can be calculated as:

$$\beta = \left(\frac{I}{\lambda} + H^T H \right)^{-1} H^T T' \tag{13}$$

When the number of concealed layer nodes exceeds the number of training data, β is⁶⁴:

$$\beta = H^T \left(\frac{I}{\lambda} + H^T H \right)^{-1 T} T' \tag{14}$$

Support vector regression (SVR). Smola and Scholkopf⁶⁶ have written a detailed description of SVR methodology. As a result, this study mainly provides a basic overview of SVR. SVR fits the linear relationship $y = x^T p + b$ regarding the designation error reduction theory, in which the regression parameters p and b are determined by minimizing the given cost function R ⁶⁷:

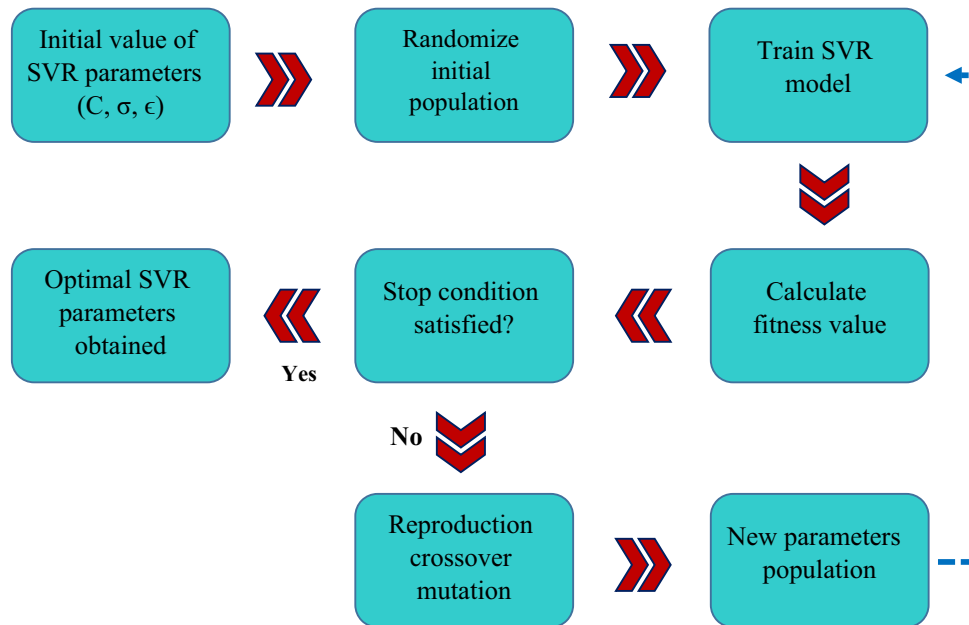


Figure 4. Flowchart of the SVR model.

$$R = \frac{1}{2}P^T P + C \frac{1}{J} \sum_{j=1}^J L_\epsilon(d_j, y_j) \tag{15}$$

$$L_\epsilon(d_j, y_j) = \begin{cases} |d_j, y_j| - \epsilon & |d_j, y_j| \geq \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{16}$$

here, d_j and y_j refer to the measured and predicted values of the j th training component, respectively, and J is the count of training components. This cost function aims to minimize both model intricacy ($1/2(P^T P)$) and actual error ($C \frac{1}{J} \sum_{j=1}^J L_\epsilon(d_j, y_j)$) at the same time. C is a regularization coefficient that controls the model intricacy vs. actual error trade-off. $L_\epsilon(d_j, y_j)$ is the ϵ -insensitive loss function, as stated in Eq. (16). The cost function R is limited in the following scenario when slack parameters ξ are introduced⁶⁷:

$$R = \frac{1}{2}P^T P + C \sum_{j=1}^J (\xi_j, \xi_j^*), \rightarrow \begin{cases} d_j - x_j^T P - b \leq \epsilon + \xi_j \\ -d_j + x_j^T P + b \leq \epsilon + \xi_j^* \\ \xi_j^*, \xi_j \geq 0 \end{cases} \tag{17}$$

Therefore, by employing the optimization criteria and inserting Lagrange multipliers (α_j^*, α_j) linear relationship $y = x^T p + b$ has the given explicit expression⁶⁷:

$$y = \sum_{j=1}^J (\alpha_j - \alpha_j^*) (x_j, x) + b \tag{18}$$

By replacing a kernel function $K(x_j, x)$ for the dot product (x_j, x) , this linear approach may be extended to nonlinear regression in a direct way. As the kernel function, a variety of functions can be employed. The Gaussian radial basis function $K(x_j, x) = \exp(-\|x_j - x\| / (2\gamma^2))$ is a regularly utilized kernel function⁶⁷. Figure 4 shows the flowchart of the SVR approach.

Boosting support vector regression (B-SVR). For the regression problem, many implementations of boosting were utilized, including adaptive boosting for regression⁶⁸ and Stochastic gradient boosting⁶⁹. The essential objective of B-SVR is to train a series of SVR systems, each of which is built often using data with high mistakes gained from the prior SVR designs. The following is a description of the technique⁶⁷.

To begin, all of the data in the initial training dataset are given identical weights $w_j^{(1)} = 1, j = 1, \dots, J$. Then, for each computation phase $t = 1, \dots, T$ (where T is the group size or repetition number), do the foregoing: (First Stage): The probability of drawing the i th input for SVR analysis is⁶⁷:

$$f_j^{(t)} = \frac{w_j^{(t)}}{\sum_{j=1}^J w_j^{(t)}}, \quad j = 1, \dots, J. \quad (19)$$

J samples (sometimes with repetition, i.e., some values that occur several times in the boosting collection) are chosen from the main training samples according to the selection probability functions $f(t)$.

(Second stage): Utilizing boosting dataset, create an SVR network and predict the outputs of the main training data.

(Third stage): Using the given loss function, calculate a loss value for each item in the initial training dataset⁶⁷:

$$L_j = \frac{|y_j^{(t)} - d_j|}{\max(y^{(t)} - d)}, \quad j = 1, \dots, J. \quad (20)$$

(Fourth stage): Determine the mean loss:

$$\bar{L}_t = \sum_{j=1}^J L_j f_j \quad (21)$$

(Fifth stage): Consider $\beta_t = \frac{\bar{L}_t}{1 - \bar{L}_t}$ and modify the weight of each component in the previous training dataset utilizing⁶⁷:

$$w_j^{(t+1)} = w_j^{(t)} \beta_t^{(1-L_j)} \quad (22)$$

The notation $(1 - L_j)$ indicates that β_t has been increased to the $(1 - L_j)$ th power. This weight-updating technique indicates that if the SVR system generated in the second stage fails to estimate data, its weight will be raised. β_t is a measure of the developed SVR model's reliability. A high β_t suggests that the SVR system has a poor level of certainty. Because the quantity of $w_j^{(t+1)}$ is required to calculate $f_j^{(t+1)}$ in the first stage of the next computation phase $(t + 1)$, Eq. (22) yields a transition to the next calculation cycle $(t + 1)$. T is chosen by the training collection's root mean squared error (RMSE)⁷⁰, and boosting cycle is performed by Drucker⁷¹ when the average loss is less than 0.5⁶⁷.

Finally, in terms of estimation, each of the T SVR systems produces a forecasting $y_j^{(t)}$ for the j th unknown feature and a matching β_t . These T forecasts are merged to provide a target variable utilizing the weighted median method, which is calculated as follows:

Arrange these T forecasts, $y_j^{(t)}$ ($t = 1, \dots, T$), in increasing sequence for the j th sample⁶⁷:

$$y_j^{(n_1)} \leq y_j^{(n_2)} \leq \dots \leq y_j^{(n_J)} \quad (23)$$

here, n_j ($j = 1, 2, \dots, J$; $J = T$) is the initial cycle number t for the j th sample's associated forecast. Consider the case where the biggest predicted value of y_j was anticipated by the third computation cycle ($t = 3$), in which case $y_j^{(n_J)}$ should equal $y_j^{(3)}$. From the lowest term n_1 to the r th term n_r , add the sum of $\log(\frac{1}{\beta_{n_j}})$ through j until the following inequality is satisfied⁶⁷:

$$\sum_{j=1}^r \log\left(\frac{1}{\beta_{n_j}}\right) \geq \frac{1}{2} \sum_{j=1}^J \log\left(\frac{1}{\beta_{n_j}}\right) \quad (24)$$

The overall estimate for the j th data is then extracted from the forecast from the n_r th SVR system⁷¹. Utilizing the same technique as before, group estimates of outputs for additional inputs may also be produced⁶⁷.

Deep belief network (DBN). Artificial Neural Networks (ANNs) have been initially suggested in the 1960s and now have grown in prominence in a variety of machine learning approaches, including classification, clustering, regression, and forecasting⁷². ANNs are flexible analytical approaches that can build complicated associations between real-world information. Among the several types of ANNs, the Back-Propagation (BP) mechanism is one of the most common structures. Although, it should be highlighted that using casual weights at the beginning of the training process is one of the difficulties with BP. As a consequence of this challenge, a unique technique, Deep Belief Network (DBN), for deep network pre-training was devised in 2006⁷³, leading to significant advances in machine learning.

A DBN is comprised of several Restricted Boltzmann Machines (RBMs), which are unattended creative prediction model that uses one hidden layer to mimic the presumed distribution of measurable variables. Our DBN uses a series of RBMs to identify relevant architectures, which extract high-level properties from original data. Figure 5 shows a DBN in graphical form⁷⁴.

DBN training is divided into three phases⁷⁵:

1. Conduct unsupervised and hungry layer-wise pre-training utilizing a collection of RBMs.
2. After haphazardly setting the linking weights matrix between the last hidden layer and the output neuron, calculate the error (first fine-tuning stage).

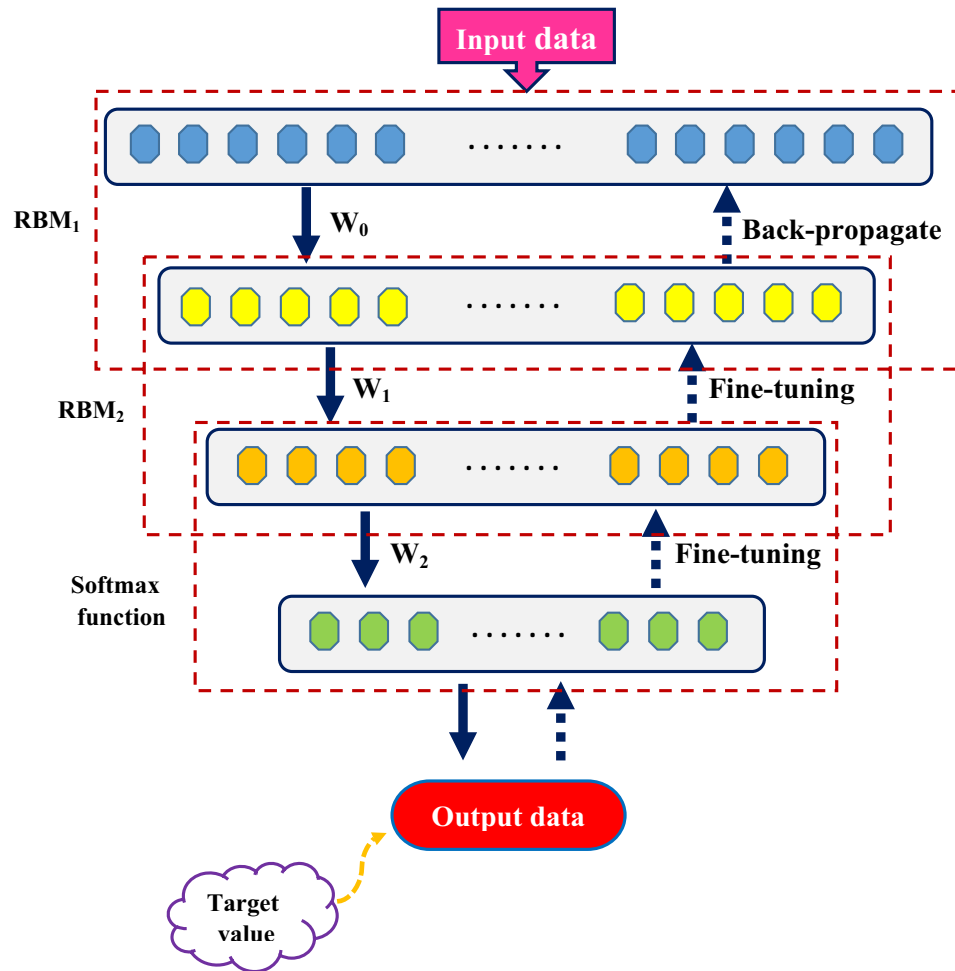


Figure 5. The DBN algorithm flowchart.

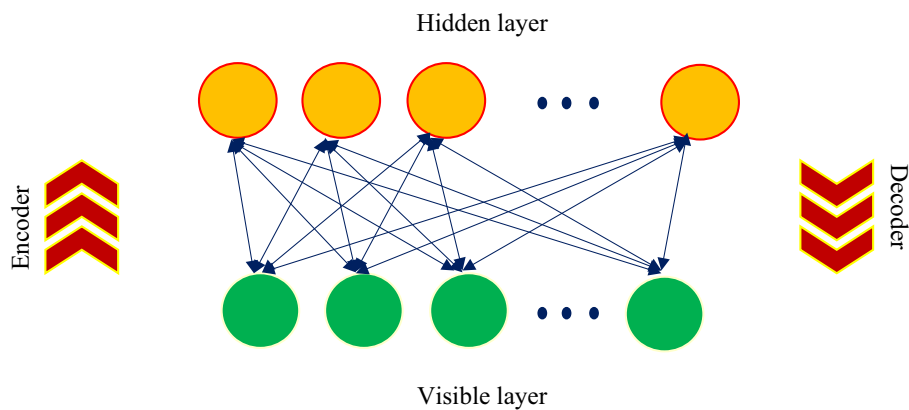


Figure 6. Schematic of the restricted Boltzmann machine.

- As a second fine-tuning stage, use error Back Propagation. The RBM training technique is described as follows:

Restricted Boltzmann machine. The encoder/decoder technique used by the Restricted Boltzmann Machine turns inputs into an increased attributes statement (Fig. 6). The decoder may then retrieve the inputs. RBM growth through the reproduction of inputs is a major feature of DBN because this strategy is unregulated and does not need marked data⁷⁶.

EOS	Reference	Relationship	Coefficients
PR3: 3-parameter Peng–Robinson	88	$P = \frac{RT}{v-b} - \frac{a}{v^2+uv+w}$	$a = \Omega_a \frac{R^2 T_c^2}{P_c}$ $\Omega_a = \Omega_{ac} \alpha(T_r)$ $b = \Omega_b \frac{RT_c}{P_c}$ $C = \Omega_c \frac{RT_c}{P_c}$ $u = \Omega_u \frac{RT_c}{P_c}$ $w = \Omega_w \left(\frac{RT_c}{P_c} \right)^2$ $\alpha = [1 + m(1 - T_r^{0.5})]^2$
SRK3: 3-parameter Soave–Redlich–Kwong	89	$P = \frac{RT}{v-b} - \frac{aa}{v(v+b)}$	$a = \frac{0.42747R^2 T_c^2}{P_c}$ $b = \frac{0.08664RT_c}{P_c}$ $\alpha = [1 + m(1 - T_r^{0.5})]^2$ $m = 0.48508 + 1.5517\omega - 0.1561\omega^2$
RK: Redlich–Kwong	90	$P = \frac{RT}{v-b} - \frac{a}{\sqrt{T}v(v+b)}$	$a = \frac{0.42748R^2 T_c^{2.5}}{P_c}$ $b = \frac{0.08664RT_c}{P_c}$
ZJ: Zudkevitch–Joffe	91	$P = \frac{RT}{v-b} - \frac{a}{\sqrt{T}v(v+b)}$	Parameter a and b are calculated as a functions of temperature and pressure

Table 4. Equations of state applied in this work.

The RBM is made up of a number of visible, $v \in \{0,1\}^{g_v}$, and concealed units, $h \in \{0,1\}^{g_h}$, the number of which is represented by the g_v and g_h , respectively. The energy of the joint link $\{v,h\}$ in RBM, considering bias, is⁷⁷:

$$E(v,h) = -v^T W h - b^T v - c^T h \quad (25)$$

The vectors of visible units, hidden units, visible unit's bias, and hidden unit's bias are denoted by x , h , b , and c , respectively. The transpose vectors of v , b , and c are denoted by v^T , b^T , and c^T , respectively. The weight matrix is also W . The following formula is used to compute the probability of each possible situation $\{v,h\}$ ⁷⁶:

$$P(v,h) = \frac{1}{Z} \exp(-E(v,h)) \quad (26)$$

here, Z is the normalizing factor and denotes:

$$Z = \sum_{v,h} \exp(-E(v,h)) \quad (27)$$

In this work, we have used intelligent models based on two different methods. The reason for choosing the thermodynamic properties-based method is to make a model whose inputs are the same as the equations of state so that it is simple and a reasonable comparison can be made. The reason for choosing the method based on chemical structure is to make a model that is more general and can be used for different ionic liquids and operating conditions.

Equations of state (EOSs)

Equations of state (EOSs) are utilized to determine the thermodynamic characteristics of pure substances and mixtures, particularly for processes at pressures higher than 10 bar, since in these circumstances at least one constituent exists in supercritical phase in most cases. The cubic EOSs are the most common types of EOS. The van der Waals EOS, proposed in 1873⁷⁸ as the first model to represent features of both liquid and vapor phases, is the source of these approaches. Numerous EOS have been built on top of the van der Waals EOS in recent decades. For instance, there are over 220 versions and numerous researchers that are relevant to parameterization and application to compounds for the widely used Peng–Robinson EOS. Cubic EOS has become quite widespread in the oil and gas industry because they are straightforwardly formulated mathematically and provide good thermodynamic property relationships and forecasts for compounds of different fluids⁷⁹. In this study, to evaluate the performance of intelligent techniques, the obtained results are compared with equations of state. In this regard, four equations of state including Peng–Robinson (PR), Soave–Redlich–Kwong (SRK), Redlich–Kwong (RK), and Zudkevitch–Joffe (ZJ) have been used. The mathematical relations of the equations mentioned are summarized in Table 4.

Results and discussion

Statistical analysis. In this study, the solubility of gaseous hydrocarbons in ILs was predicted using 2145 laboratory data with the four intelligent models already mentioned. For this purpose, two different methods based on thermodynamic properties and the chemical structure of ILs and hydrocarbons have been used. The process of model development and estimation of the solubility values is illustrated in Fig. 7. Statistical criteria

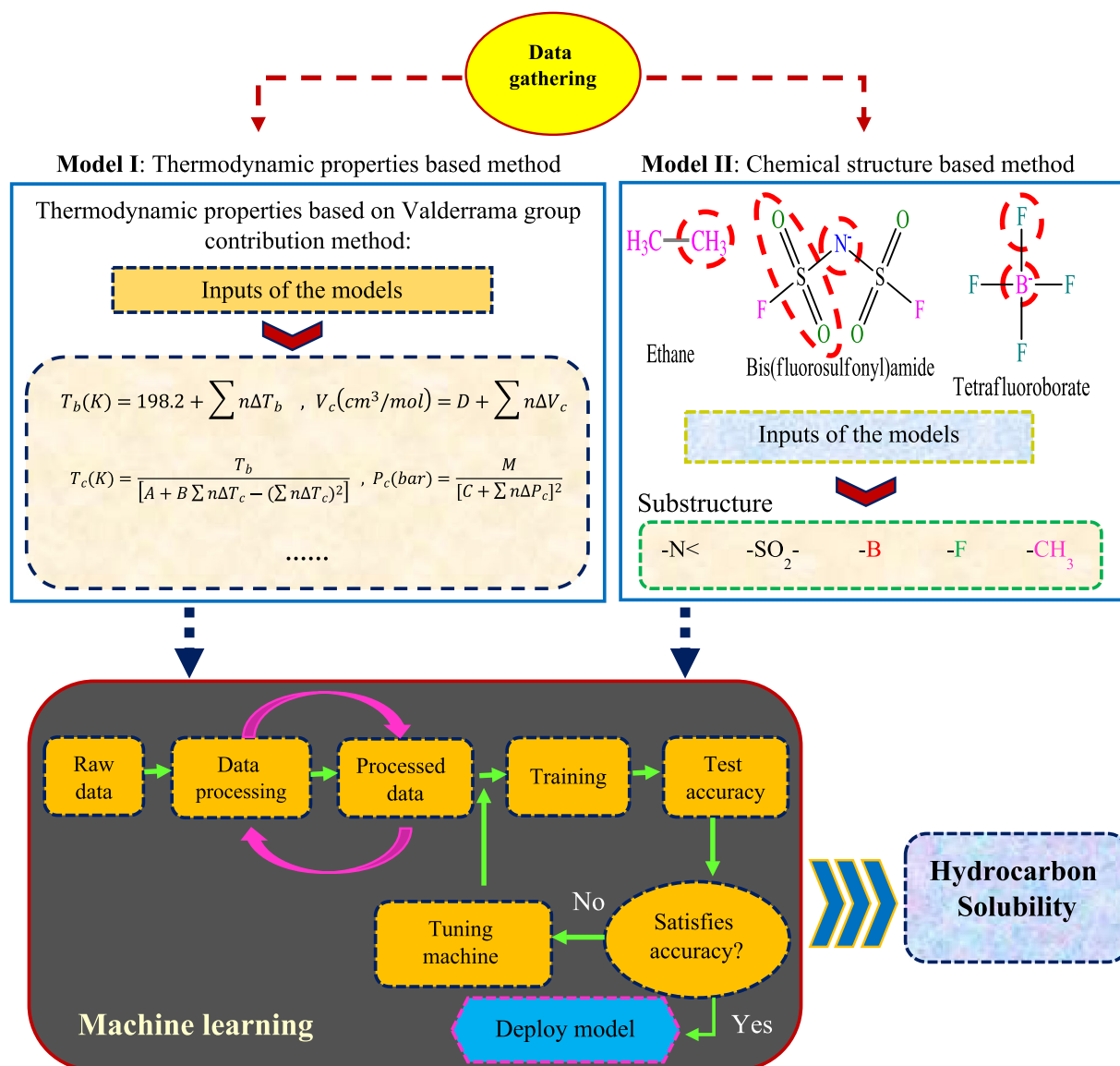


Figure 7. Procedure of data preparation and model development for hydrocarbon solubility prediction.

comparing the performance of intelligent models are root mean squared error (RMSE) and coefficient of determination (R^2), the mathematical relationships of which are defined as follows^{24,80}:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (HS_{iexp} - HS_{ipred})^2} \quad (28)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (HS_{iexp} - HS_{ipred})^2}{\sum_{i=1}^n (HS_{iexp} - \overline{HS}_{iexp})^2} \quad (29)$$

where HS_{iexp} , HS_{ipred} , \overline{HS}_{iexp} , and n refer to experimental hydrocarbon solubility, predicted hydrocarbon solubility, mean value of experimental hydrocarbon solubility, and total number of data points, respectively.

The results of statistical analysis of the developed models are reported in Table 5. According to the obtained results, the DBN model has the best performance in both methods. In the method based on thermodynamic properties, the DBN model has the highest accuracy with R^2 and RMSE values of 0.9961 and 0.0054, respectively. The R^2 and RMSE values for the DBN model in the chemical structure-based method are 0.9943 and 0.0065, respectively. Generally, the developed models are arranged in the following order of increasing accuracy:

Model (I): MARS < B-SVR < ELM < DBN.

Model (II): MARS < B-SVR < ELM < DBN.

Model	Data	Statistical parameter			
		Model (I)		Model (II)	
		RMSE	R ²	RMSE	R ²
Boosting-SVR	Train	0.0083	0.9906	0.0075	0.9923
	Test	0.0119	0.9825	0.0110	0.9849
	Total	0.0091	0.9889	0.0083	0.9907
DBN	Train	0.0033	0.9985	0.0055	0.9959
	Test	0.0101	0.9874	0.0097	0.9883
	Total	0.0054	0.9961	0.0065	0.9943
ELM	Train	0.0052	0.9963	0.0052	0.9964
	Test	0.0105	0.9863	0.0111	0.9846
	Total	0.0066	0.9942	0.0068	0.9938
MARS	Train	0.0135	0.9752	0.0104	0.9851
	Test	0.0169	0.9646	0.0152	0.9714
	Total	0.0142	0.9729	0.0115	0.9822

Table 5. Statistical analysis results of developed models.

Graphical analysis. Plotting the predicted values versus the experimental values obtain a cross plot diagram. In this diagram, the more points that correspond to the $X = Y$ line, the closer the predicted values are to the actual values, so the accuracy of the model will increase. Figure 8 shows a cross-plot diagram for models developed based on thermodynamic properties. As can be seen from Fig. 8, the DBN model has higher density points around the $X = Y$ line, which indicates the higher accuracy of this model than other models, although other models are also in an acceptable situation. Figure 9 shows the cross-plot diagram of the results of the method based on chemical structure, temperature, and pressure. The graphical analysis presented in Fig. 9 also confirms the results of Table 5, and the DBN model has a more appropriate cross plot diagram than the other models.

Figures 10 and 11 illustrate the distribution of the deviation of the predicted values from the actual values for Model (I) and Model (II), respectively. The higher the focus around the zero line, the closer the predicted values are to the actual values and the more accurate the model. Therefore, as can be seen from Figs. 10 and 11, the DBN model is in a better position than the other smart models in both methods.

The Taylor diagram⁸¹ was developed in Fig. 12, which integrates a number of statistical variables for a more understandable representation. This illustration depicts all of the smart approaches in respect of how well they forecast gaseous hydrocarbon solubility in various ILs. Three performance parameters [standard deviation (SD), RMSE, and R^2] of each network, including DBN, ELM, B-SVR, and MARS, are utilized to quantify the degree of discrepancy between the systems' estimates and the related actual values. The reference point is a distance from the full measure that is used to calculate the centered RMSE. Another reference is the optimal estimation technique, which is denoted by a point with R^2 equal to 1. Figure 12a,b show the Taylor diagram of the DBN model for the results of Model (I) and Model (II), respectively. In Fig. 12a, the RMSE and R^2 values are 0.0054 and 0.9961, respectively. In Fig. 12b, the RMSE and R^2 values are 0.0065 and 0.9943, respectively. Figure 12 depicts the performance parameters of the systems after they have been encapsulated within the Taylor diagram. The Taylor diagram in this picture verifies DBN's dominance by demonstrating that their worldwide performance assessment is the most accurate.

Equations of state evaluation. Equations of state are one of the conventional methods for calculating the solubility of gases in ILs. Comparing the results of smart models with equations of state can reveal the efficiency and performance of smart models well. For this purpose, four equations of state including Peng–Robinson (PR), Soave–Redlich–Kwong (SRK), Redlich–Kwong (RK), and Zudkevitch–Joffe (ZJ) have been used to calculate the solubility of gaseous hydrocarbons in ILs. Figure 13 shows the results obtained for two systems ([BMIM][PF₆]-CH₄ and [BMIM][PF₆]-C₂H₆) at 323.15 K. As can be seen from Fig. 13, with increasing pressure, the solubility of gaseous hydrocarbons increases, a trend that can also be seen in the equations of state. Also, the solubility of C₂H₆ in [BMIM][PF₆] is higher compared to CH₄. Another noteworthy point is the very good agreement of the results of the smart models with the experimental data. It can also be found that the equations of state overestimate the solubility and the difference between the calculated values and the experimental data is also significant.

Hamedi et al.⁵ calculated the solubility of methane in ILs using two equations of state including modified Sanchez and Lacombe (SL) and cubic-plus-association (CPA). In this study, 649 experimental data related to 19 ionic liquid and methane systems were used. Figure 14 shows a comparison of current study models and Hamedi et al. results. As can be seen, the accuracy of the intelligent models compared to the equations of state used by Hamedi et al.⁵ is significant. The advantages of the proposed models in the current study over the work of Hamedi et al. are the use of a wider database (2145 experimental data) including more diverse systems (48 systems) and more advanced general models. Also, the results obtained in this study are more accurate than the research of Hamedi et al.

Trend analysis. Involvement of various parameters such as temperature, pressure, intermolecular interactions, etc. in the solubility of gases in ILs makes it difficult to find physical phenomena affecting the transfer

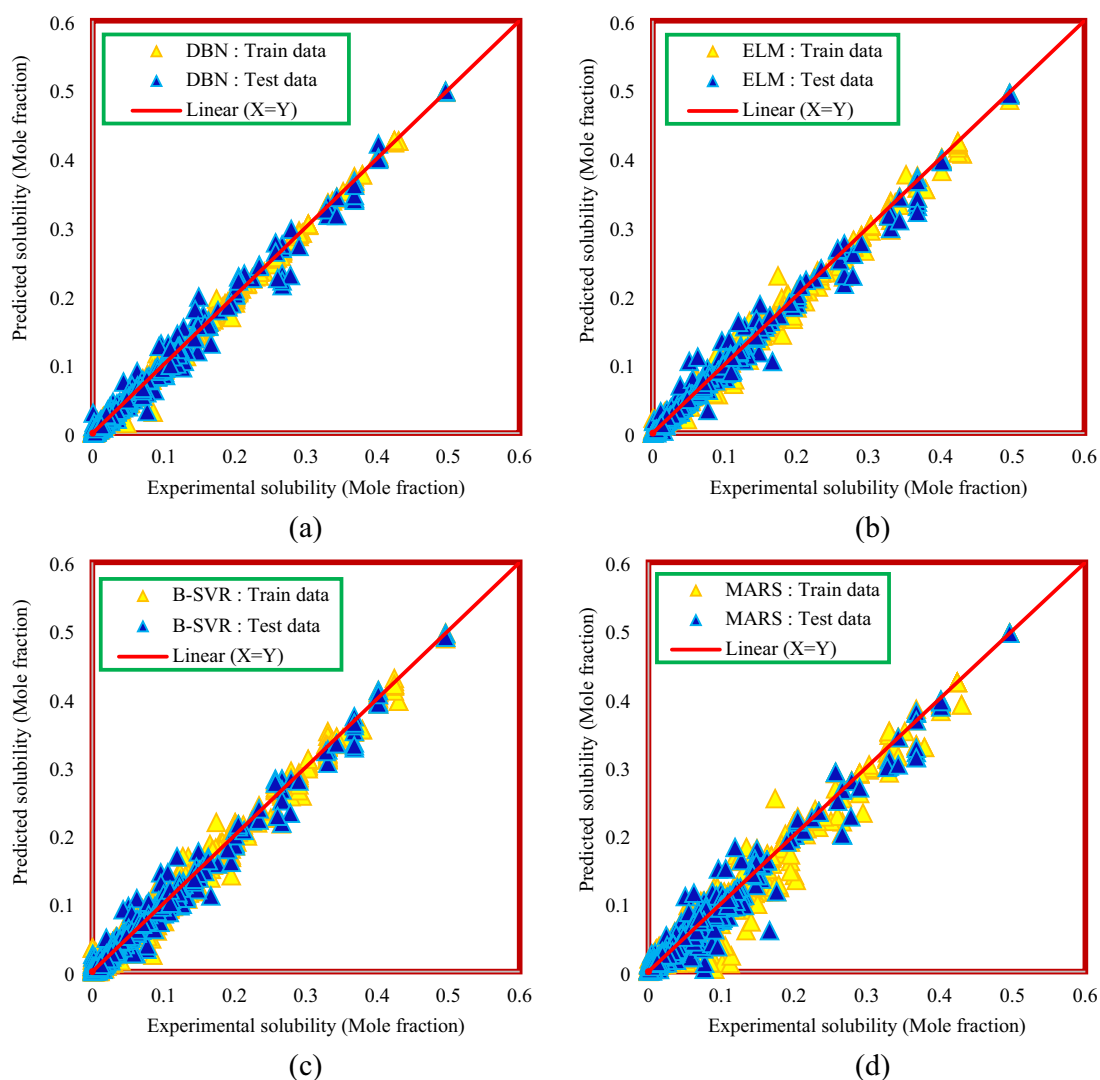


Figure 8. Cross plot diagram of the results of Model (I): (a) DBN model, (b) ELM model, (c) B-SVR model, (d) MARS model.

process, but in simpler terms, changing solubility factors (such as Henry's constant) due to changes in physical and chemical properties cause these trends^{82,83}. Enthalpy, entropy, and interactions between gas and ionic liquid are the most important factors affecting the solubility of gases in ILs. Therefore, the effect of changes in pressure, temperature, type of anion and cation on solubility is affected by changes in enthalpy and entropy. In other words, the endothermic or exothermic nature of the dissolution process, which is also depends on the interactions between the gas and the ionic liquid, affects how the gas solubility changes with the increase of each of the mentioned parameters⁸⁴.

Pressure and anion type effect. Previous studies have shown that although ILs have similar solubility at low pressures⁸⁵, increasing pressure leads to increasing the solubility of gases in ILs and the rate of increase varies for different ILs⁸⁶. Figure 15 shows the changes in the solubility of gaseous hydrocarbons with increasing pressure for a number of systems used in this study. As can be seen, increasing the pressure increases the solubility of gaseous hydrocarbons in ILs. The effect of the type of anions existing in the structure of ILs has been studied in many studies⁸⁶. As mentioned in the literature, the main reasons for changes in the solubility due to changes in pressure and anion type are enthalpy and entropy changes, and the interactions between gas and ILs⁸⁴. The effect of a number of anions has been investigated for example. Investigation of the systems reported in Fig. 15 shows that the effect of anion type on the solubility of C_2H_6 in ILs is as follows:



Cation effect. In previous studies, the influence of cation type has been investigated⁴³. The reasons for the effect of cation type on solubility are similar to the reasons mentioned for the effect of anion type. In this study, the

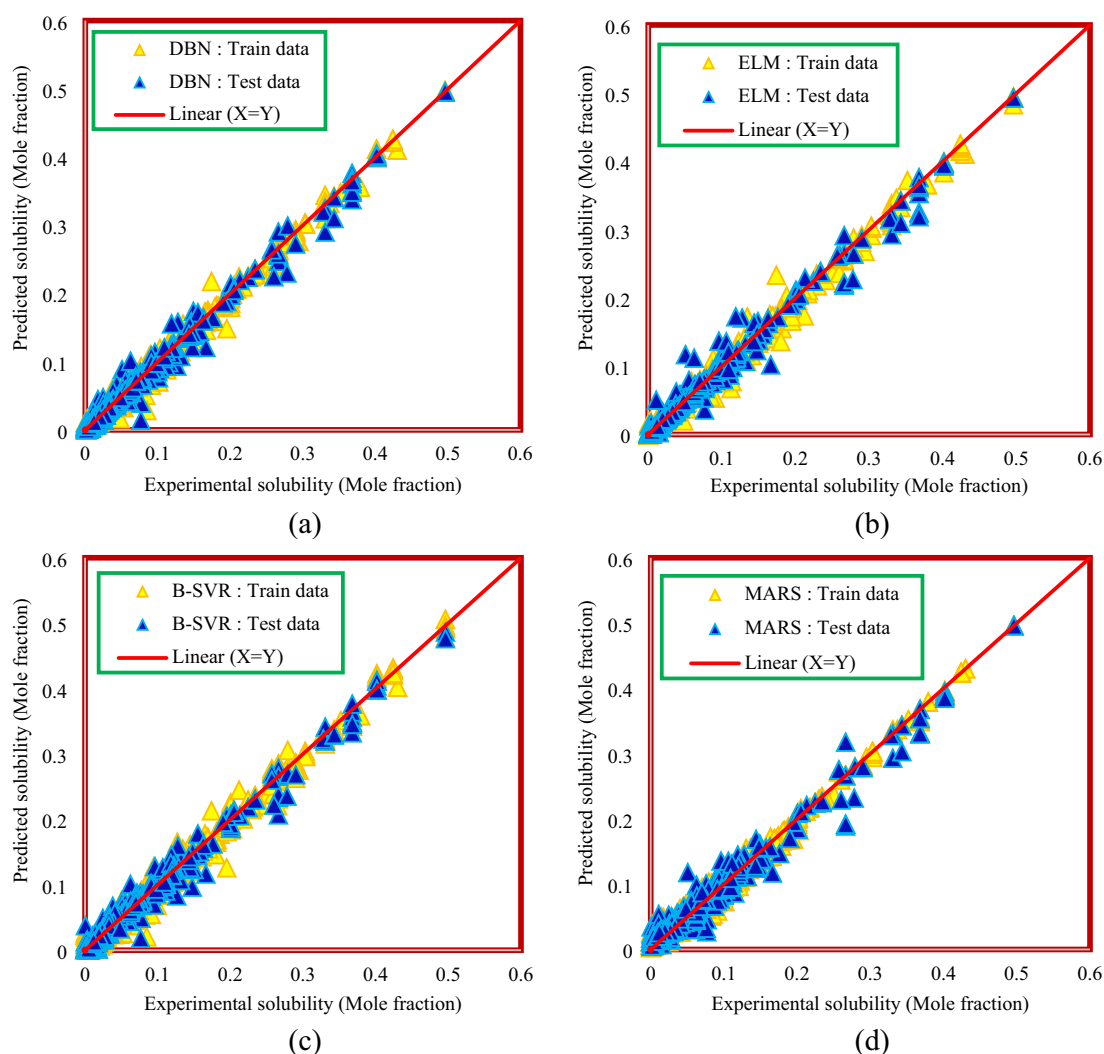
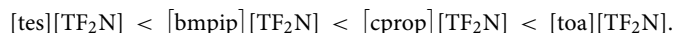


Figure 9. Cross plot diagram of the results of Model (II): (a) DBN model, (b) ELM model, (c) B-SVR model, (d) MARS model.

effect of a number of cations on the solubility of gaseous hydrocarbons in ILs has been studied. Figure 16 examines a number of systems studied in this work. According to the results obtained for the proposed systems, the effect of cation type on the solubility of methane in ILs is as follows:



Temperature and alkyl chain length effect. Many researchers have studied the effect of temperature⁸⁶ and alkyl chain length⁸⁷ on the solubility of gases in ILs. The results show that increasing the temperature reduces the solubility⁸⁶. Generally, the dependency of solubility on temperature is thermodynamically related to the enthalpy of dissolution. For example, in exothermic processes (such as dissolution of CO_2 in ILs) the solubility of the gas decreases with increasing temperature, and in endothermic processes (such as dissolution of N_2 in ILs) the increase in temperature increases the solubility^{83,84}. Also, increasing the alkyl chain length increases the solubility of hydrocarbons in ILs⁸⁷. The reason for this subject is the increased entropy of solvation for ILs⁸⁴. According to Fig. 17, although increasing the alkyl chain length does not show a specific trend in ethane solubility in the proposed systems, increasing the temperature has reduced the ethane solubility in the proposed systems.

Effect of the type of hydrocarbon. In order to investigate the effect of hydrocarbon type on the solubility of gaseous hydrocarbons, four different systems have been investigated. Figure 18 shows the trend of solubility changes for these four systems with increasing pressure at 320 K. According to the results, it can be seen that the solubility of proposed hydrocarbons in $[\text{BMIM}][\text{TF}_2\text{N}]$ is determined as follows:

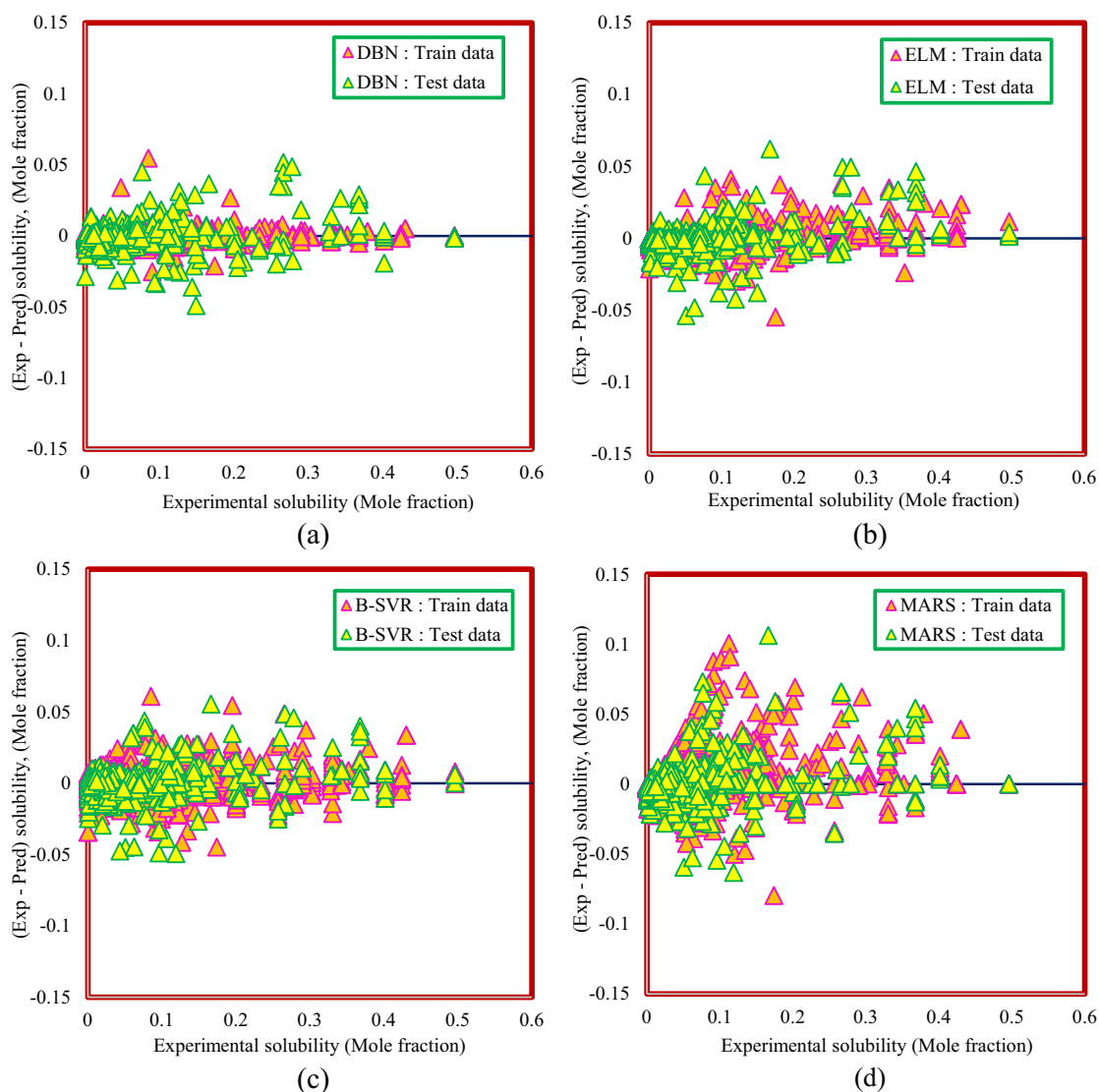
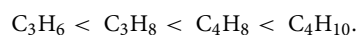


Figure 10. Distribution of deviation of predicted values from actual values for Model (I): (a) DBN model, (b) ELM model, (c) B-SVR model, (d) MARS model.



Therefore, it can be concluded that increasing the molecular weight of hydrocarbons can increase their solubility in ILs.

Therefore, as it was investigated, different parameters can affect the solubility of gaseous hydrocarbons in ionic liquids, and this effect can vary depending on the type of gas and ionic liquid. For example, Ramdin et al.⁴ reported that although the enthalpy of absorption of CO₂ is higher in [amim][dca] than in [thtdp][phos], its solubility is lower in [amim][dca]. On the other hand, this effect was the opposite for CH₄, and the solubility of methane in ionic liquids corresponded to the enthalpy of adsorption. Also, its solubility in ionic liquids with longer non-polar alkyl chains is higher. Thus, entropic effects and solvent–solvent interactions prevailed for CO₂ and CH₄, respectively. The solubility of hydrocarbons in ILs is important in many ways. The selection of ILs as solvents in reactions involving permanent gas and gas storage media is strongly influenced by the degree of gas solubility in ILs and is essential for the development of separation methods. Investigation of the effect of various parameters such as temperature, pressure, etc. is necessary to deal with the separation of CO₂ from natural gas as well as the low adsorption of light hydrocarbons in ILs such as methane, ethane and propane. Therefore, evaluating the effect of temperature, pressure, anion and cation type, hydrocarbon type, and alkyl chain length can help researchers to select the appropriate ILs in different operating conditions. The findings of the study also show that ILs are suitable candidates for the solubility of hydrocarbons so they can be used in various operational applications.

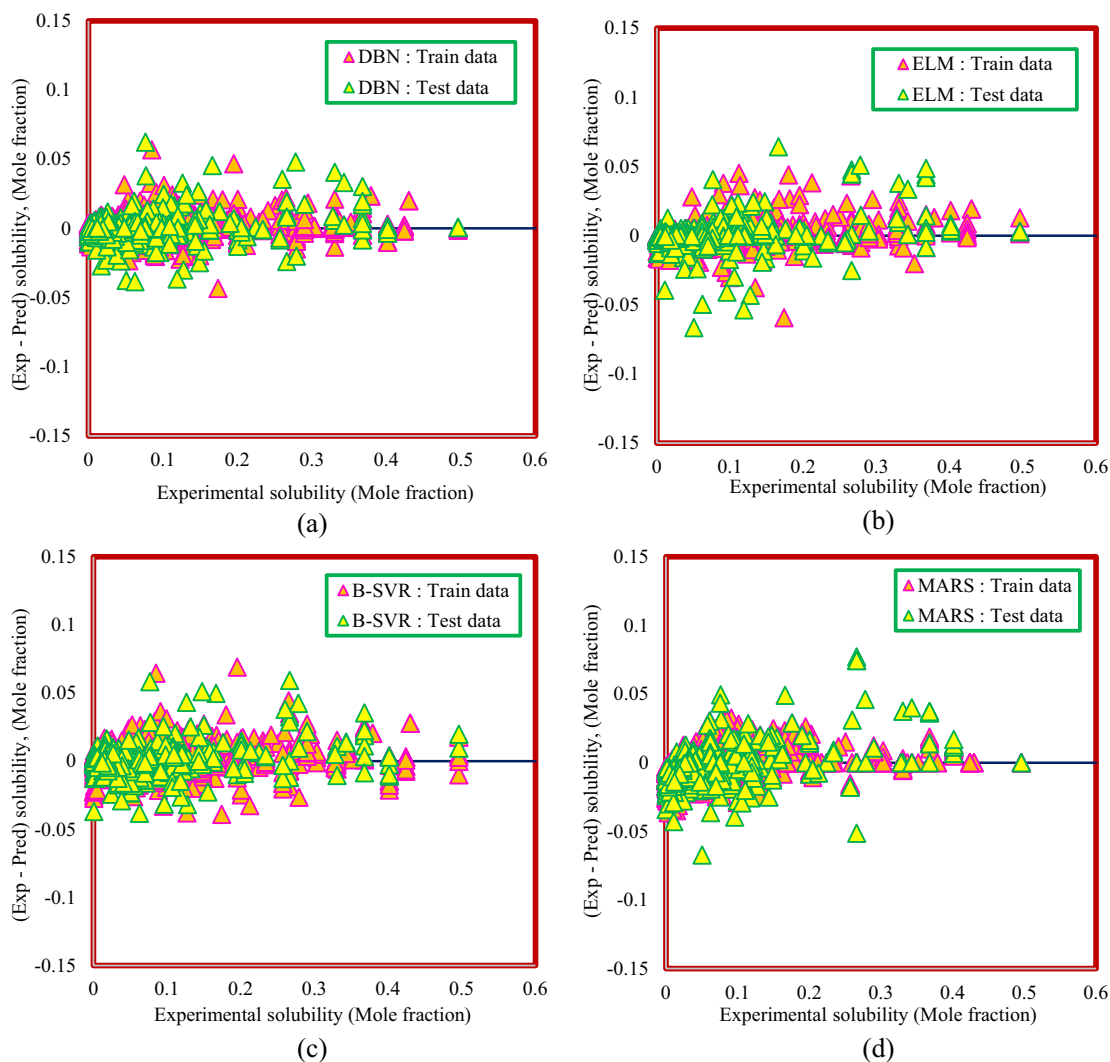


Figure 11. Distribution of deviation of predicted values from actual values for Model (II): (a) DBN model, (b) ELM model, (c) B-SVR model, (d) MARS model.

Conclusions

In this work, the solubility of gaseous hydrocarbons in ILs was predicted using DBN, ELM, Boost-SVR, and MARS models. In this regard, the thermodynamic properties and chemical substructures of gaseous hydrocarbons and ILs along with temperature and pressure have been used in distinct methods to predict the solubility of gaseous hydrocarbons in ILs. The findings of this study give the following conclusions:

- In the method based on thermodynamic properties, the DBN model with RMSE and R^2 values of 0.0054 and 0.9961, respectively, has the most accurate estimates.
- In the method based on chemical structure, temperature, and pressure, the DBN model with RMSE and R^2 values of 0.0065 and 0.9943, respectively, has the best performance.
- Comparison of the results of smart models with equations of state shows that the accuracy of the former is much higher than the latter.
- The study of the effect of different factors on the solubility of gaseous hydrocarbons shows that increasing the pressure and molecular weight of hydrocarbons increases the solubility, while increasing the temperature decreases the solubility of gaseous hydrocarbons in ILs.
- Investigation of the effect of the type of anion and cation in the structure of ILs shows that the following order have been established for the solubility of the studied systems:

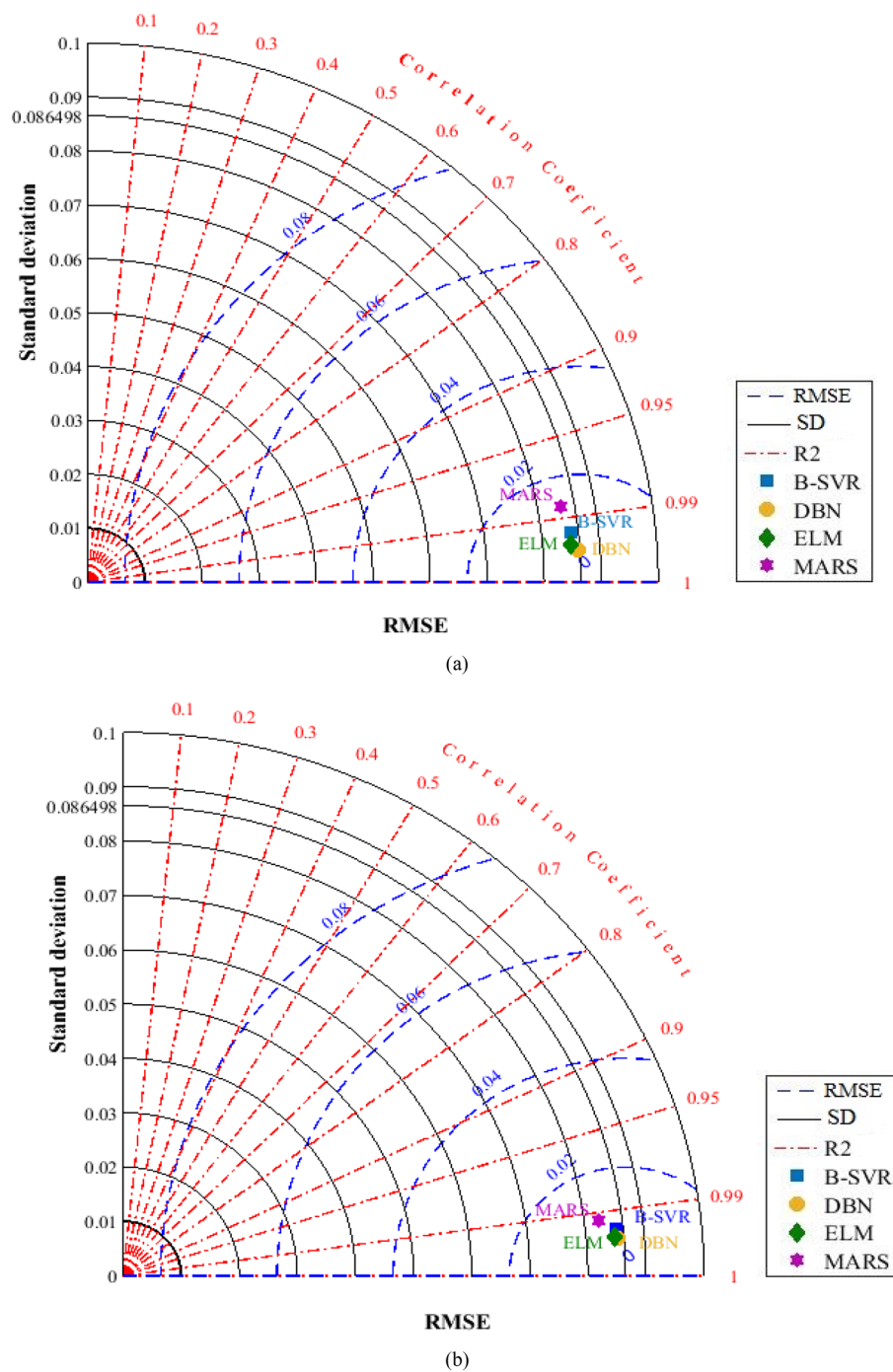
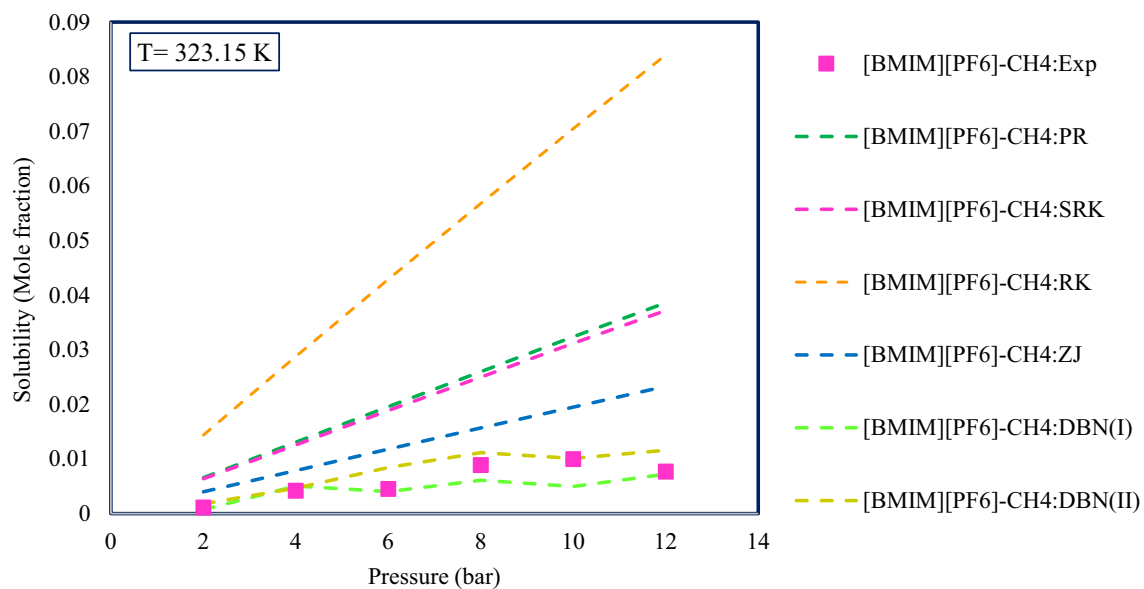


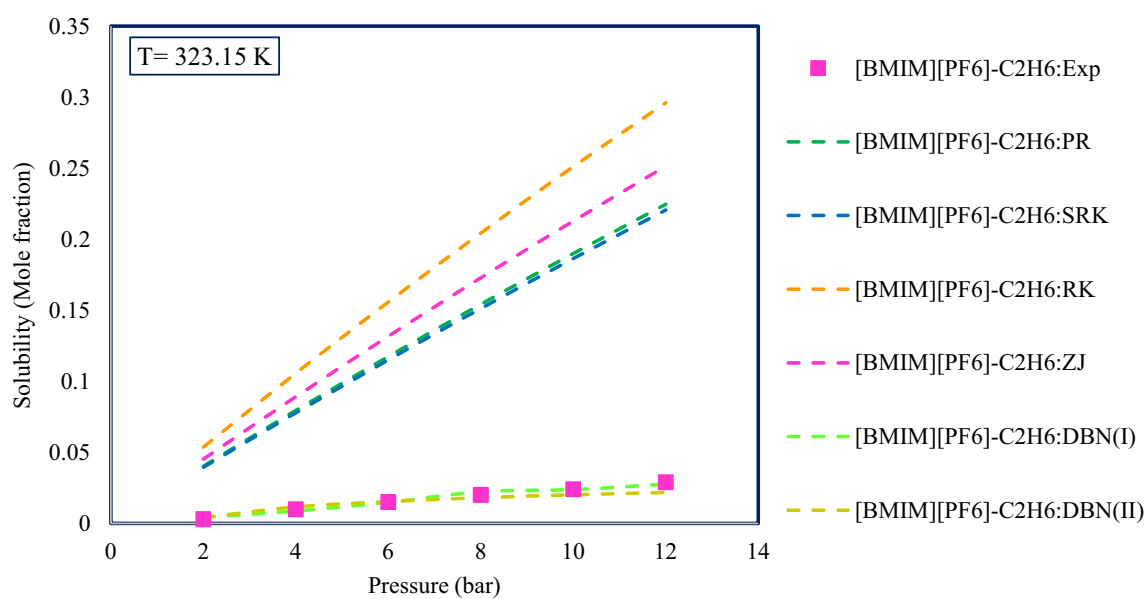
Figure 12. Taylor diagram of developed models: (a) Model (I), (b) Model (II).

Order of the solubility of C₂H₆ in ILs based on the anion type: [BMIM][PF₆] < [BMIM][TF₂N] < [BMIM][eFAP]

Order of the solubility of CH₄ in ILs based on the cation type: [tes][TF₂N] < [bmpip][TF₂N] < [cprop][TF₂N] < [toa][TF₂N]



(a)



(b)

Figure 13. Comparison of the results of the DBN models and equations of state; (a) [BMIM][PF₆]-CH₄, (b) [BMIM][PF₆]-C₂H₆.

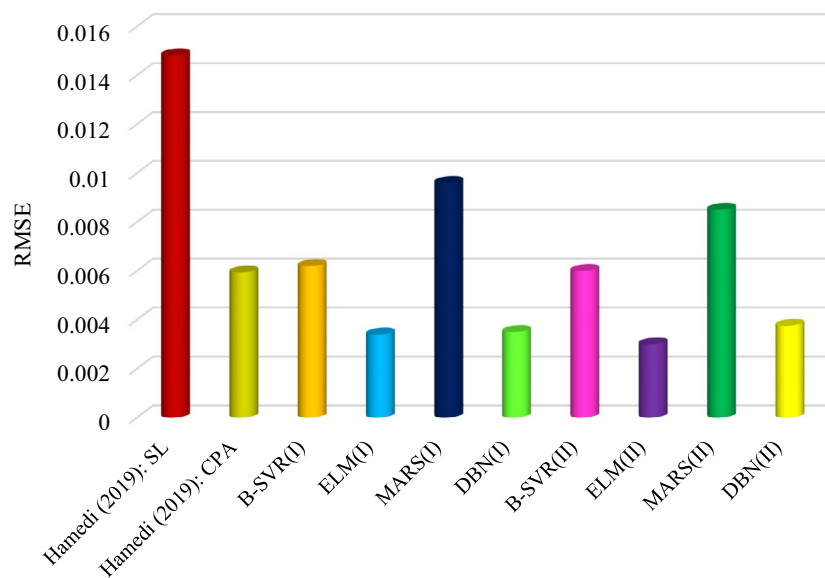


Figure 14. Comparison of the results of the current study models and equations of state used in previous studies.

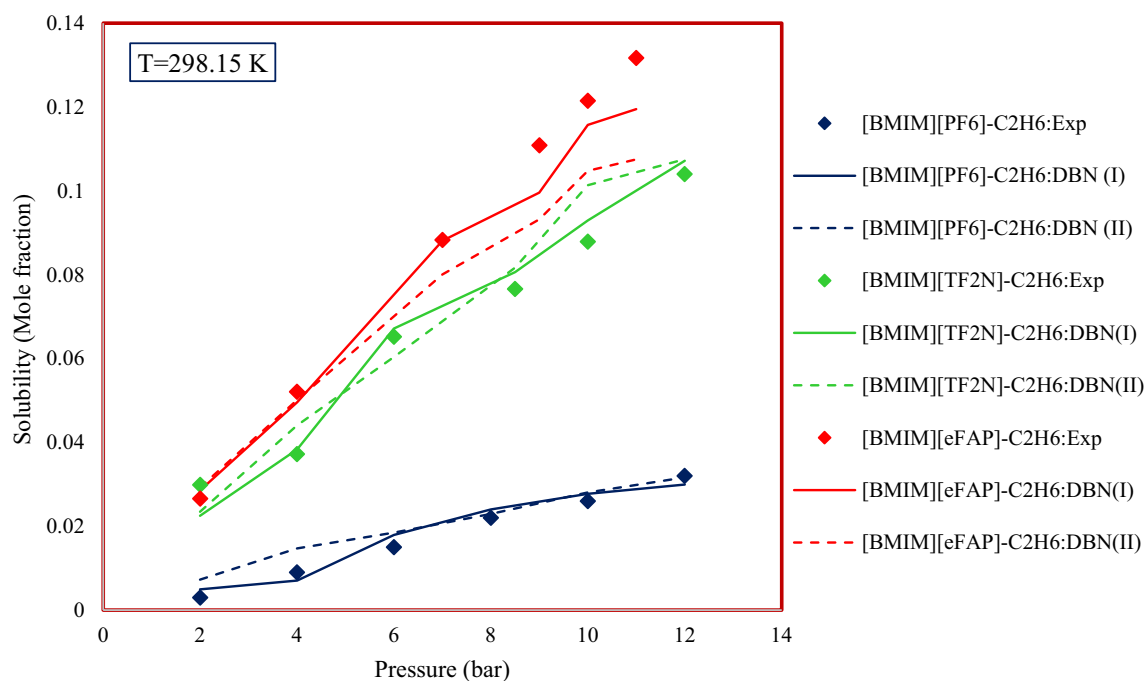


Figure 15. Pressure and anion type effect on the solubility of gaseous hydrocarbons in ionic liquids.

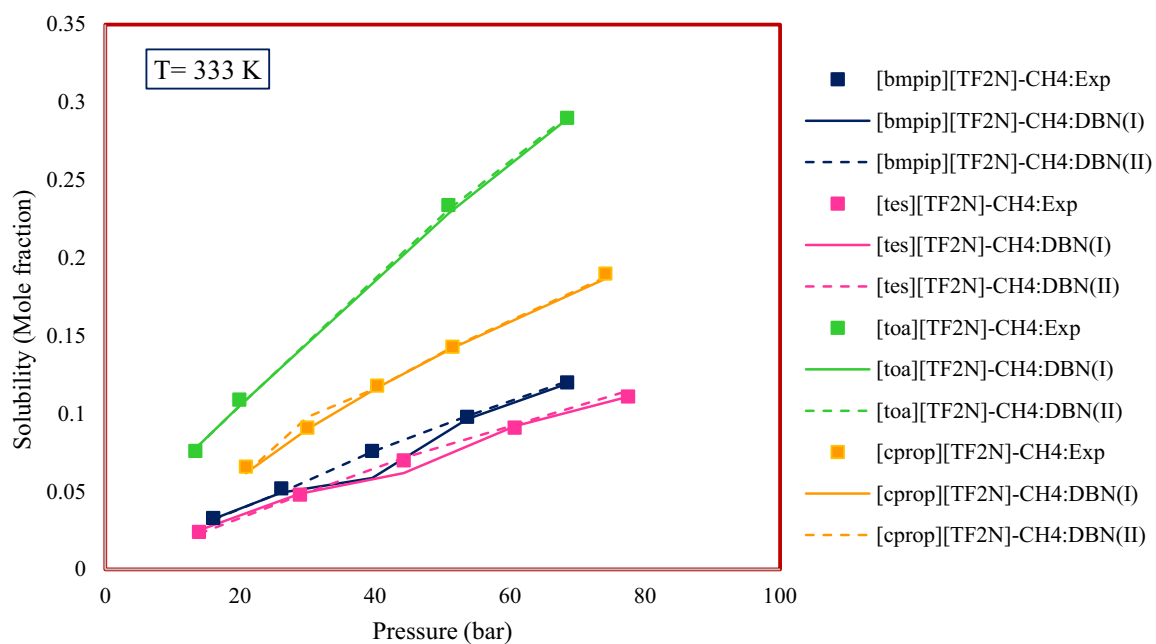


Figure 16. Cation type effect on the solubility of gaseous hydrocarbons in ionic liquids.

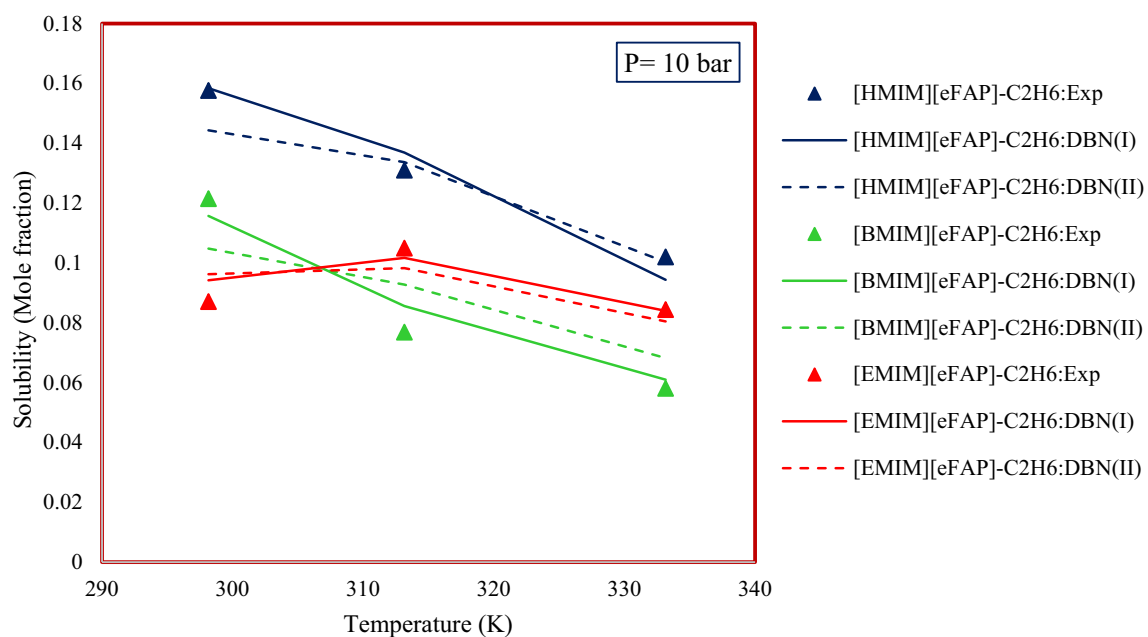


Figure 17. Temperature and alkyl chain length effect on the solubility of gaseous hydrocarbons in ionic liquids.

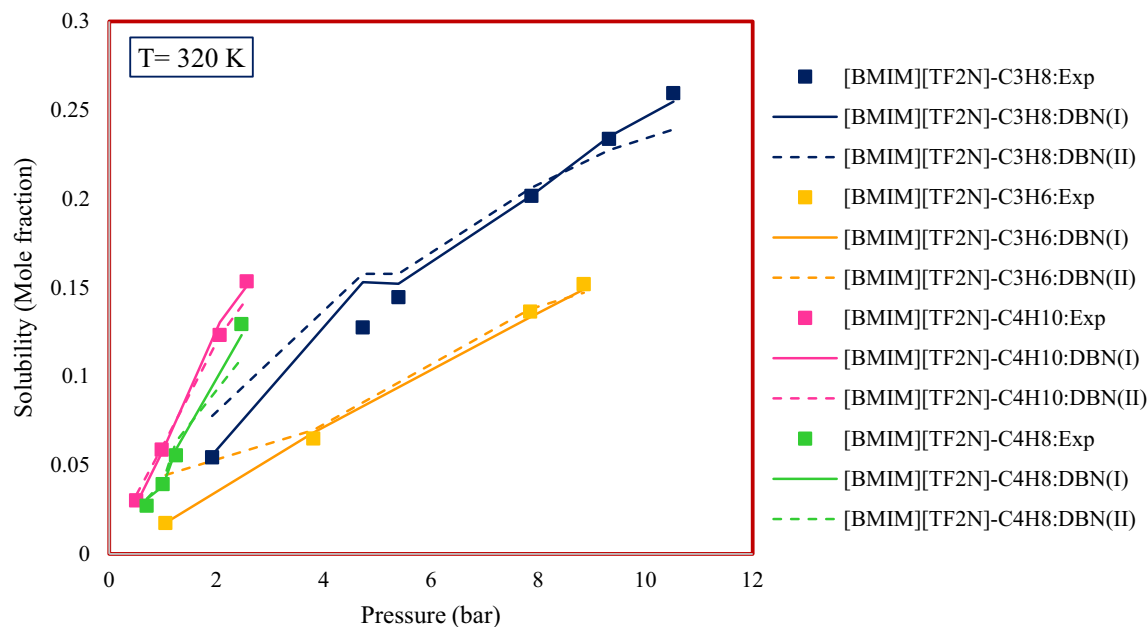


Figure 18. Type of hydrocarbon effect on the solubility of gaseous hydrocarbons in ionic liquids.

Data availability

All the data have been collected from literature. We cited all the references of the data in the manuscript. However, the data will be available from the corresponding author on reasonable request.

Received: 11 March 2022; Accepted: 3 August 2022

Published online: 22 August 2022

References

- Atashrouz, S., Zarghampour, M., Abdolrahimi, S., Pazuki, G. & Nasernejad, B. Estimation of the viscosity of ionic liquids containing binary mixtures based on the Eyring's theory and a modified Gibbs energy model. *J. Chem. Eng. Data* **59**, 3691–3704 (2014).
- Atashrouz, S., Mozaffarian, M. & Pazuki, G. Modeling the thermal conductivity of ionic liquids and ionanofluids based on a group method of data handling and modified Maxwell model. *Ind. Eng. Chem. Res.* **54**, 8600–8610 (2015).
- Raeissi, S. & Peters, C. High pressure phase behaviour of methane in 1-butyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide. *Fluid Phase Equilib.* **294**, 67–71 (2010).
- Ramdin, M. *et al.* Solubility of CO₂ and CH₄ in ionic liquids: Ideal CO₂/CH₄ selectivity. *Ind. Eng. Chem. Res.* **53**, 15427–15435 (2014).
- Hamed, N., Rahimpour, M. R. & Keshavarz, P. Methane solubility in ionic liquids: Comparison of cubic-plus-association and modified Sanchez–Lacombe equation of states. *Chem. Phys. Lett.* **738**, 136903 (2020).
- Durand, E., Lecomte, J. & Villeneuve, P. Deep eutectic solvents: Synthesis, application, and focus on lipase-catalyzed reactions. *Eur. J. Lipid Sci. Technol.* **115**, 379–385 (2013).
- Sekerci-Cetin, M., Emek, O. B., Yildiz, E. E. & Unlusu, B. Numerical study of the dissolution of carbon dioxide in an ionic liquid. *Chem. Eng. Sci.* **147**, 173–179 (2016).
- Pison, L. *et al.* Solubility of n-butane and 2-methylpropane (isobutane) in 1-alkyl-3-methylimidazolium-based ionic liquids with linear and branched alkyl side-chains. *Phys. Chem. Chem. Phys.* **17**, 30328–30342 (2015).
- Anderson, J. L., Anthony, J. L., Brennecke, J. F. & Maginn, E. J. Gas solubilities in ionic liquids. *Ion. Liq. Synth.* **1**, 103–129 (2008).
- Karadas, F., Atilhan, M. & Aparicio, S. Review on the use of ionic liquids (ILs) as alternative fluids for CO₂ capture and natural gas sweetening. *Energy Fuels* **24**, 5817–5828 (2010).
- Chen, Y., Mutelet, F. & Jaubert, J.-N. Solubility of carbon dioxide, nitrous oxide and methane in ionic liquids at pressures close to atmospheric. *Fluid Phase Equilib.* **372**, 26–33 (2014).
- Huseynov, M. *Thermodynamic and Experimental Studies of Ethane Solubility in Promising Ionic Liquids for CO₂ Capture* (The University of Regina (Canada), 2014).
- Stevanovic, S. & Gomes, M. C. Solubility of carbon dioxide, nitrous oxide, ethane, and nitrogen in 1-butyl-1-methylpyrrolidinium and trihexyl (tetradecyl) phosphonium tris (pentafluoroethyl) trifluorophosphate (eFAP) ionic liquids. *J. Chem. Thermodyn.* **59**, 65–71 (2013).
- Jalal, A., Can, E., Keskin, S., Yildirim, R. & Uzun, A. Selection rules for estimating the solubility of C₄-hydrocarbons in imidazolium ionic liquids determined by machine-learning tools. *J. Mol. Liq.* **284**, 511–521 (2019).
- Song, Z. *et al.* Solubility of imidazolium-based ionic liquids in model fuel hydrocarbons: A COSMO-RS and experimental study. *J. Mol. Liq.* **224**, 544–550 (2016).
- AlSaleem, S. S., Zahid, W. M., AlNashef, I. M. & Hadj-Kali, M. K. Solubility of halogenated hydrocarbons in hydrophobic ionic liquids: Experimental study and COSMO-RS prediction. *J. Chem. Eng. Data* **60**, 2926–2936 (2015).
- Panayiotou, C. & Sanchez, I. Hydrogen bonding in fluids: An equation-of-state approach. *J. Phys. Chem.* **95**, 10090–10097 (1991).
- Koutsoukos, S., Philippi, F., Malaret, F. & Welton, T. A review on machine learning algorithms for the ionic liquid chemical space. *Chem. Sci.* **12**, 6820–6843 (2021).
- Shahriari, S., Atashrouz, S. & Pazuki, G. Mathematical model of the phase diagrams of ionic liquids-based aqueous two-phase systems using the group method of data handling and artificial neural networks. *Theor. Found. Chem. Eng.* **52**, 146–155 (2018).

20. Mousavi, S. P. *et al.* Viscosity of ionic liquids: Application of the Eyring's theory and a committee machine intelligent system. *Molecules* **26**, 156 (2021).
21. Mousavi, S.-P. *et al.* Modeling surface tension of ionic liquids by chemical structure-intelligence based models. *J. Mol. Liq.* **342**, 116961 (2021).
22. Ding, Y., Chen, M., Guo, C., Zhang, P. & Wang, J. Molecular fingerprint-based machine learning assisted QSAR model development for prediction of ionic liquid properties. *J. Mol. Liq.* **326**, 115212 (2021).
23. Mousavi, S. P. *et al.* Modeling thermal conductivity of ionic liquids: A comparison between chemical structure and thermodynamic properties-based models. *J. Mol. Liq.* **322**, 114911 (2021).
24. Mousavi, S. P. *et al.* Modeling of H₂S solubility in ionic liquids using deep learning: A chemical structure-based approach. *J. Mol. Liq.* **351**, 118418 (2021).
25. Ferreira, A. R. *et al.* An overview of the liquid-liquid equilibria of (ionic liquid + hydrocarbon) binary systems and their modeling by the conductor-like screening model for real solvents. *Ind. Eng. Chem. Res.* **50**, 5279–5294 (2011).
26. Shaahmadi, F., Anbaz, M. A. & Bazooyar, B. Analysis of intelligent models in prediction nitrous oxide (N₂O) solubility in ionic liquids (ILs). *J. Mol. Liq.* **246**, 48–57 (2017).
27. Bahmani, A. R., Sabzi, F. & Bahmani, M. Prediction of solubility of sulfur dioxide in ionic liquids using artificial neural network. *J. Mol. Liq.* **211**, 395–400 (2015).
28. Mokarizadeh, H., Atashrouz, S., Mirshekar, H., Hemmati-Sarapardeh, A. & Pour, A. M. Comparison of LSSVM model results with artificial neural network model for determination of the solubility of SO₂ in ionic liquids. *J. Mol. Liq.* **304**, 112771 (2020).
29. Faúndez, C. A., Fierro, E. N. & Valderrama, J. O. Solubility of hydrogen sulfide in ionic liquids for gas removal processes using artificial neural networks. *J. Environ. Chem. Eng.* **4**, 211–218 (2016).
30. Xia, L., Wang, J., Liu, S., Li, Z. & Pan, H. Prediction of CO₂ solubility in ionic liquids based on multi-model fusion method. *Processes* **7**, 258 (2019).
31. Baghban, A., Ahmadi, M. A. & Shahraki, B. H. Prediction carbon dioxide solubility in presence of various ionic liquids using computational intelligence approaches. *J. Supercrit. Fluids* **98**, 50–64 (2015).
32. Shafiei, A. *et al.* Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach. *J. Supercrit. Fluids* **95**, 525–534 (2014).
33. Song, Z., Shi, H., Zhang, X. & Zhou, T. Prediction of CO₂ solubility in ionic liquids using machine learning methods. *Chem. Eng. Sci.* **223**, 115752 (2020).
34. Moosanezhad-Kermani, H., Rezaei, F., Hemmati-Sarapardeh, A., Band, S. S. & Mosavi, A. Modeling of carbon dioxide solubility in ionic liquids based on group method of data handling. *Eng. Appl. Comput. Fluid Mech.* **15**, 23–42 (2021).
35. Mohammadi, M.-R. *et al.* Toward predicting SO₂ solubility in ionic liquids utilizing soft computing approaches and equations of state. *J. Taiwan Inst. Chem. Eng.* **133**, 104220. <https://doi.org/10.1016/j.jtice.2022.104220> (2022).
36. Alcantara, M. L. *et al.* High pressure vapor-liquid equilibria for binary protic ionic liquids + methane or carbon dioxide. *Sep. Purif. Technol.* **196**, 32–40 (2018).
37. Althuluth, M., Kroon, M. C. & Peters, C. J. High pressure solubility of methane in the ionic liquid 1-hexyl-3-methylimidazolium tricyanomethanide. *J. Supercrit. Fluids* **128**, 145–148 (2017).
38. Anderson, J. L., Dixon, J. K. & Brennecke, J. F. Solubility of CO₂, CH₄, C₂H₆, C₂H₄, O₂, and N₂ in 1-Hexyl-3-methylpyridinium Bis (trifluoromethylsulfonyl) imide: Comparison to other ionic liquids. *Acc. Chem. Res.* **40**, 1208–1216 (2007).
39. Mirzaei, M., Mokhtarani, B., Badii, A. & Sharifi, A. Solubility of carbon dioxide and methane in 1-hexyl-3-methylimidazolium nitrate ionic liquid, experimental and thermodynamic modeling. *J. Chem. Thermodyn.* **122**, 31–37 (2018).
40. Oliveira, L. M. *et al.* High pressure vapor-liquid equilibria for binary methane and protic ionic liquid based on propionate anions. *Fluid Phase Equilib.* **426**, 65–74 (2016).
41. Anthony, J. L., Anderson, J. L., Maginn, E. J. & Brennecke, J. F. Anion effects on gas solubility in ionic liquids. *J. Phys. Chem. B* **109**, 6366–6374 (2005).
42. Anthony, J. L., Maginn, E. J. & Brennecke, J. F. Solubilities and thermodynamic properties of gases in the ionic liquid 1-n-butyl-3-methylimidazolium hexafluorophosphate. *J. Phys. Chem. B* **106**, 7315–7320 (2002).
43. Hong, G. *et al.* Solubility of carbon dioxide and ethane in three ionic liquids based on the bis (trifluoromethyl) sulfonyl imide anion. *Fluid Phase Equilib.* **257**, 27–34 (2007).
44. Jacquemin, J., Gomes, M. F. C., Husson, P. & Majer, V. Solubility of carbon dioxide, ethane, methane, oxygen, nitrogen, hydrogen, argon, and carbon monoxide in 1-butyl-3-methylimidazolium tetrafluoroborate between temperatures 283 K and 343 K and at pressures close to atmospheric. *J. Chem. Thermodyn.* **38**, 490–502 (2006).
45. Almantariotis, D. *et al.* Absorption of carbon dioxide, nitrous oxide, ethane and nitrogen by 1-alkyl-3-methylimidazolium (C n mim, n = 2, 4, 6) tris (pentafluoroethyl) trifluorophosphate ionic liquids (eFAP). *J. Phys. Chem. B* **116**, 7728–7738 (2012).
46. Bermejo, M. D., Fieback, T. M. & Martín, Á. Solubility of gases in 1-alkyl-3-methylimidazolium alkyl sulfate ionic liquids: Experimental determination and modeling. *J. Chem. Thermodyn.* **58**, 237–244 (2013).
47. Costa Gomes, M. Low-pressure solubility and thermodynamics of solvation of carbon dioxide, ethane, and hydrogen in 1-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl) amide between temperatures of 283 K and 343 K. *J. Chem. Eng. Data* **52**, 472–475 (2007).
48. Kumeľan, J., Pérez-Salado Kamps, Á., Tuma, D. & Maurer, G. Solubility of the single gases methane and xenon in the ionic liquid [bmim][CH₃SO₃]. *J. Chem. Eng. Data* **52**, 2319–2324 (2007).
49. Yuan, X. *et al.* Solubilities of gases in 1, 1, 3, 3-tetramethylguanidium lactate at elevated pressures. *J. Chem. Eng. Data* **51**, 645–647 (2006).
50. Kumeľan, J., Pérez-Salado Kamps, Á., Tuma, D. & Maurer, G. Solubility of the single gases methane and xenon in the ionic liquid [hmim][TF₂N]. *Ind. Eng. Chem. Res.* **46**, 8236–8240 (2007).
51. Althuluth, M., Kroon, M. C. & Peters, C. J. Solubility of methane in the ionic liquid 1-ethyl-3-methylimidazolium tris (pentafluoroethyl) trifluorophosphate. *Ind. Eng. Chem. Res.* **51**, 16709–16712 (2012).
52. Florusse, L. J., Raeissi, S. & Peters, C. J. High-pressure phase behavior of ethane with 1-hexyl-3-methylimidazolium bis (trifluoromethylsulfonyl) imide. *J. Chem. Eng. Data* **53**, 1283–1285 (2008).
53. Kim, Y., Jang, J., Lim, B., Kang, J. W. & Lee, C. Solubility of mixed gases containing carbon dioxide in ionic liquids: Measurements and predictions. *Fluid Phase Equilib.* **256**, 70–74 (2007).
54. Jacquemin, J., Husson, P., Majer, V. & Gomes, M. F. C. Low-pressure solubilities and thermodynamics of solvation of eight gases in 1-butyl-3-methylimidazolium hexafluorophosphate. *Fluid Phase Equilib.* **240**, 87–95 (2006).
55. Geng, J., Li, M.-W., Dong, Z.-H. & Liao, Y.-S. Port throughput forecasting by MARS-RSVR with chaotic simulated annealing particle swarm optimization algorithm. *Neurocomputing* **147**, 239–250 (2015).
56. Friedman, J. H., & Roosen, C. B. An introduction to multivariate adaptive regression splines. *Stat. Methods Med. Res.* **4**(3), 197–217 (1995).
57. Zhang, W. & Goh, A. T. C. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Comput. Geotech.* **48**, 82–95 (2013).
58. Li, D. H., Chen, W., Li, S. & Lou, S. Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS)—A case study of Hong Kong. *Energy* **186**, 115857 (2019).

59. Maleki, A. *et al.* Thermal conductivity modeling of nanofluids with ZnO particles by using approaches based on artificial neural network and MARS. *J. Therm. Anal. Calorim.* **143**, 4261–4272 (2021).
60. Fridedman, J. Multivariate adaptive regression splines (with discussion). *Ann. Stat.* **19**, 79–141 (1991).
61. Zhang, W. & Goh, A. T. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geosci. Front.* **7**, 45–52 (2016).
62. Huang, G.-B. An insight into extreme learning machines: Random neurons, random features and kernels. *Cogn. Comput.* **6**, 376–390 (2014).
63. Huang, G.-B., Zhou, H., Ding, X. & Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **42**, 513–529 (2011).
64. Xiao, D., Li, B. & Mao, Y. A multiple hidden layers extreme learning machine method and its application. *Math. Probl. Eng.* **2017**, 1–10 (2017).
65. Wei, J., Liu, H., Yan, G. & Sun, F. Robotic grasping recognition using multi-modal deep extreme learning machine. *Multidimens. Syst. Signal Process.* **28**, 817–833 (2017).
66. Smola, A. J. & Schölkopf, B. A tutorial on support vector regression. *Stat. Comput.* **14**, 199–222 (2004).
67. Zhou, Y.-P. *et al.* Boosting support vector regression in QSAR studies of bioactivities of chemical compounds. *Eur. J. Pharm. Sci.* **28**, 344–353 (2006).
68. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).
69. Friedman, J. H. Stochastic gradient boosting. *Comput. Stat. Data Anal.* **38**, 367–378 (2002).
70. He, P., Xu, C.-J., Liang, Y.-Z. & Fang, K.-T. Improving the classification accuracy in chemistry via boosting technique. *Chemom. Intell. Lab. Syst.* **70**, 39–46 (2004).
71. Drucker, H. Improving regressors using boosting techniques. *ICML* **97**, 107–115 (1997).
72. Ahmadizar, F., Soltanian, K., AkhlaghianTab, F. & Tsoulos, I. Artificial neural network development by means of a novel combination of grammatical evolution and genetic algorithm. *Eng. Appl. Artif. Intell.* **39**, 1–13 (2015).
73. Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
74. Ren, J., Green, M. & Huang, X. In *Learning Control* 205–219 (Elsevier, 2021).
75. Zheng, S. & Zhao, J. In *Computer Aided Chemical Engineering* vol. 44, 2239–2244 (Elsevier, 2018).
76. Yang, X.-S. *Introduction to Algorithms for Data Mining and Machine Learning* (Academic Press, 2019).
77. Bengio, Y. *Learning Deep Architectures for AI* (Now Publishers Inc, 2009).
78. Van der Waals, J. D. *Over de Continuïteit van den Gas-en Vloeistoftoestand*. vol. 1 (Sijthoff, 1873).
79. Papadopoulos, A. I., Tsivintzelis, I., Linke, P. & Seferlis, P. *Computer Aided Molecular Design: Fundamentals, Methods and Applications* (2018).
80. Shahabi-Ghahfarokhy, A., Nakhaei-Kohani, R., Amar, M. N. & Hemmati-Sarapardeh, A. Modelling density of pure and binary mixtures of normal alkanes: Comparison of hybrid soft computing techniques, gene expression programming, and equations of state. *J. Petrol. Sci. Eng.* **208**, 109737 (2022).
81. Hu, Z., Chen, X., Zhou, Q., Chen, D. & Li, J. DISO: A rethink of Taylor diagram. *Int. J. Climatol.* **39**, 2825–2832 (2019).
82. Camper, D., Scovazzo, P., Koval, C. & Noble, R. Gas solubilities in room-temperature ionic liquids. *Ind. Eng. Chem. Res.* **43**, 3049–3054 (2004).
83. Liu, H., Dai, S. & Jiang, D.-E. Solubility of gases in a common ionic liquid from molecular dynamics based free energy calculations. *J. Phys. Chem. B* **118**, 2719–2725 (2014).
84. Moura, L. *Ionic liquids for the separation of gaseous hydrocarbons*, PhD Thesis (2014).
85. Anthony, J. L. *Gas Solubilities in Ionic Liquids: Experimental Measurements and Applications*. (University of Notre Dame, 2004).
86. Galán Sánchez, L. Functionalized ionic liquids: Absorption solvents for carbon dioxide and olefin separation (2008).
87. Ramdin, M., de Loos, T. W. & Vlucht, T. J. State-of-the-art of CO₂ capture with ionic liquids. *Ind. Eng. Chem. Res.* **51**, 8149–8177 (2012).
88. Danesh, A., Xu, D.-H. & Todd, A. Comparative study of cubic equations of state for predicting phase behaviour and volumetric properties of injection gas-reservoir oil systems. *Fluid Phase Equilib.* **63**, 259–278 (1991).
89. Elsharkawy, A. M. Predicting the dew point pressure for gas condensate reservoirs: Empirical models and equations of state. *Fluid Phase Equilib.* **193**, 147–165 (2002).
90. Pedersen, K. S., Christensen, P. L., Shaikh, J. A. & Christensen, P. L. *Phase Behavior of Petroleum Reservoir Fluids* (CRC Press, 2006).
91. Ronze, D., Fongarland, P., Pitault, I. & Forissier, M. Hydrogen solubility in straight run gasoil. *Chem. Eng. Sci.* **57**, 547–553 (2002).

Author contributions

R.N.-K.: Investigation, data curation, visualization, writing-original draft, S.A.: Writing-review & editing, methodology, validation, F.H.: Conceptualization, validation, modeling, A.B.: Writing-review & editing, validation, A.H.-S.: Methodology, validation, supervision, writing-review & editing. A.M.: Writing-review & editing, validation.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.A., A.H.-S. or A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022