Separation Science and Technology

ISSN: 0149-6395 (Print) 1520-5754 (Online) Journal homepage: https://www.tandfonline.com/loi/lsst20

Modeling CO_2 absorption in aqueous solutions of DEA, MDEA, and DEA + MDEA based on intelligent methods

Danial Abooali, Reza Soleimani & Ali Rezaei-Yazdi

To cite this article: Danial Abooali, Reza Soleimani & Ali Rezaei-Yazdi (2019): Modeling CO_2 absorption in aqueous solutions of DEA, MDEA, and DEA + MDEA based on intelligent methods, Separation Science and Technology, DOI: 10.1080/01496395.2019.1575415

To link to this article: https://doi.org/10.1080/01496395.2019.1575415

SEPARATION SCIENCE AND TECHNOLOGY
https://doi.org/10.1080/01496395.2019.1575415

Taylor & Francis
Taylor & Francis Group

Modeling CO_2 absorption in aqueous solutions of DEA, MDEA, and DEA + MDEA based on intelligent methods

Danial Abooali[a], Reza Soleimani[b], and Ali Rezaei-Yazdi[c]

[a]Young Researchers and Elite Club, Central Tehran Branch, Islamic Azad University, Tehran, Iran; [b]Department of Chemical Engineering, Tarbiat Modares University, Tehran, Iran; [c]Engineering & Applied Science School, Aston University, Birmingham, UK

ABSTRACT

Removing CO_2 as an acidic-potential component from different gaseous flows is a main topic in different industries producing green-house gases, especially in natural gas sweetening units. A group of well-known absorbents for CO_2 are the amine solutions. The common amine compounds consisting di-ethanolamine (DEA), methyl-di-ethanolamine (MDEA), and their mixture in aqueous solution have been investigated in this study. The effort was to develop new models for estimation of CO_2 loading capacity of the presented amine solutions using genetic programing (GP) and stochastic gradient boosting (SGB) trees as two advanced and novel machine learning approaches in this area. A total of 175 sets of experimental data of CO_2 absorption including independent variables (temperature, CO_2 partial pressure, concentrations of DEA and MDEA in water) and objective function (CO_2 loading capacity) were collected from literature and fed to the mentioned algorithms (GP and SGB) as input dataset. Then, each algorithm was run over the dataset, separately and two new models were created. Finally, strict statistical evaluations were implemented to assess the estimating capability of the new models. The statistical parameters including correlation coefficients (R^2_{SGB} = 0.99848 and R^2_{GP} = 0.99087), root-mean-square deviations ($RMSD_{SGB}$ = 0.00903 mol/mol and $RMSD_{GP}$ = 0.02244 mol/mol) and average absolute relative deviations ($AARD_{SGB}$ = 0.95628% and $AARD_{GP}$ = 8.71909%) show that the utilized powerful algorithms have enhanced the applicability of the new developed models providing good` estimations in operational processes. Final results show superiority and more accuracy of the new SGB model for confident predictions in amine process.

ARTICLE HISTORY
Received 21 February 2018
Accepted 24 January 2019

KEYWORDS
Carbon dioxide; di-ethanolamine; methyl-di-ethanolamine; stochastic gradient boosting; genetic programming

Introduction

Production of carbon dioxide (CO_2) as a result of various industrial processes and burning large quantities of fossil fuels is one of the main environmental issues. CO_2 is present in many natural and industrial gaseous sources and various industrial cases need to remove this acidic-potential component from gas mixtures to reduce its green-house harmful effects. In different industrial units such as ammonia production, natural gas processing, removing CO_2 and other acid gases is carried out by use of amine solutions in which amine components absorb/capture acid gas molecules due to their alkaline nature. This is a well-known reliable technique which has wide applications in related industries.

There are a number of amine compounds applied in CO_2 removal processes involving the primary amines (*i.e.* the amine components including one organic group connected to nitrogen ($R\text{-}NH_2$)), secondary amines (the components including two organic groups connected to nitrogen ($R_2\text{-}NH$)), and tertiary amines (the amines with three organic groups ($R_3\text{-}N$)).[1] Mono-ethanol amine (MEA), di-ethanol amine (DEA), Methyl,di-ethanol amine (MDEA), tri-ethanol amine (TEA), piperazine (PZ), 2-amino,2-methyl-1-propanol (AMP), and di-isopropanol amine (DIPA) are applicable often-used components in industrial applications.

The mechanism of amine process for CO_2 capture involves physical absorption and chemical reactions in which both physical and chemical equilibria are determinative. Amine and CO_2 exhibit alkaline and acidic behaviors, respectively, and subsequently, they tend to make a reaction with each other. Each amine takes part in the reaction in a particular manner. But generally, similar steps take place in the reaction of all types of amine including protonation of amine and ionization of some species in the solution.[2] The efficiency of CO_2 absorption process is usually deduced by equilibrium CO_2 loading capacity of solution (α). It is an important

CONTACT Danial Abooali danial.abooali@gmail.com Young Researchers and Elite Club, Central Tehran Branch, Islamic Azad University, Tehran, Iran
Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/lsst.
Supplemental data for this article can be accessed here.

parameter which is calculated as amount (usually moles) of absorbed CO_2 per one mole of amine component. The general mass balance calculations of CO_2 absorption in amine solutions is conducted by the following relations[3]:

$$G + S = E + R \quad (1)$$

$$Gy_1 + Sx_1 = Ey_2 + Rx_2 \quad (2)$$

In which, G is mass (or mol) rate of gas feed, S is mass (or mol) rate of liquid solvent (primary amine solution), E is mass (or mol) rate of extract phase (*i.e.* the gas flow with non-absorbed CO_2), and R is mass (or mol) rate of raffinate phase (*i.e.* the liquid solution with absorbed CO_2). x_i and y_i are mass (or mol) fraction of CO_2 in liquid and gas phase, respectively. x_1 is mass (or mol) fraction of CO_2 in the solvent phase which is usually zero. x_2 is mass (or mol) fraction of CO_2 in raffinate. y_1 and y_2 are mass (or mol) fractions of CO_2 in the gas feed and extract phase, respectively.

Reactions of primary and secondary amines lead to form carbamate ion. In this case, the process is implemented in a manner with generally low CO_2 loadings capacity and high rate of absorption. However, tertiary amines exhibit opposite behavior because they do not form carbamate ions.[2] The mechanisms of these reactions are available in the literature.[1,2,4] Briefly, the reactions of amine process for CO_2 absorption are presented as follows:

Dissociation of CO_2:	$CO_2 + 2H_2O \rightleftharpoons HCO_3^- + H_3O^+$	(3)
Dissociation of HCO_3^- :	$HCO_3^- + H_2O \rightleftharpoons CO_3^{2-} + H_3O^+$	(4)
Ionization of water:	$2H_2O \rightleftharpoons OH^- + H_3O^+$	(5)
Dissociation of protonated amine of n^{th} type:	$R_nN^+H_{3-n+1} + H_2O \rightleftharpoons R_nNH_{3-n} + H_3O^+$	(6)
Formation of carbamate from primary amines:	$RNH_2 + HCO_3^- \rightleftharpoons RNHCOO^- + H_2O$	(7)
Formation of carbamate from secondary amines:	$RR'NH + HCO_3^- \rightleftharpoons RR'NCOO^- + H_2O$	(8)
Reaction of tertiary amines:	$RR'R''N + H_2O \rightleftharpoons RR'R''NH^+ + OH^-$ $CO_2 + OH^- \rightleftharpoons HCO_3^-$	(9) (10)

Choosing a certain amine (single or mixture) for CO_2 capture mainly depends on the loading capacity, reaction kinetics, components regenerating potential, and facilities.[5] Loading capacity (α) often considered as the most important factor[6] is the criterion to investigate absorbent effectiveness. Some variables affect CO_2 loading capacity during amine process: temperature (T), pressure of CO_2 gas (P_{CO2}), and concentration of amine components (C_{Amine}).

Access to loading capacity data in various operating conditions is essential for different industrial and theoretical studies about CO_2 absorption. Real values are calculated by experimental measurements in laboratories. However, in some cases, especially in high difficult conditions, it is not simple to run experiments. In some situations, experimental measurement is expensive and/or time-consuming or involves uncertainties about impurities, etc. Application of estimating methods and mathematical models can be very effective in this area. For estimation of loading capacity, thermodynamic calculation methods are available. Moreover, machine learning approaches such as genetic programming (GP), artificial neural networks (ANNs), generalized regression neural networks (GRNs), adaptive neuro-fuzzy inference system (ANFIS), support vector machines (SVMs), particle swarm optimization (PSO), and imperialist competitive algorithm (ICA) could be successfully used for developing data-based models as appropriate alternative methods to conventional thermodynamic models.[7–17]

Literature review

Amine process for CO_2 absorption has been investigated in different modeling studies. In many project, different ANN-based methods have been taken into account. In 2004, an ANN model has been developed by Daneshvar et al.[18] for CO_2 capturing in aqueous solutions of tri-isopropanol amine (TIPA), MEA, and TIPA + MEA. Sipöcz et al.[19] presented a feed-forward ANN model for CO_2 absorption on MEA from data of a power plant flue gas.

Shahsavand et al.[20] applied a pair of approaches including multi-layer perceptron neural network and radial basis function neural network in order to estimate CO_2 loading capacity for DEA and MDEA solutions. Another ANN-based model along with Deshmukh-Mather model was suggested and compared by Pahlavanzadeh et al.[21] for solubility of CO_2 in 2-amino-2-methyl-1-propanol (AMP) solution. ANN along with ANFIS approach has been applied to produce model for amine-based CO_2 loading capacity in some researches.[22,23] Least squares support vector machine (LSSVM) was also used to estimate CO_2 loading capacity of MEA, DEA, and TEA by Ghiasi and Mohammadi.[14]

In a new study, Saghafi et al.[16] proposed a model for CO_2 loading capacity of DEA + MDEA solution based on random forest (RF) technique and compared with modified Deshmukh–Mather method developed by Benamor and Aroua.[2] They applied 88 experimental data proposed by Benamor and Aroua[2] to develop the RF model and their results show more accuracy of the RF-based model.

As an interesting investigation, comparison between applicability of RF) technique and powerful methodologies of genetic programming (GP) and stochastic gradient boosting (SGB) trees for common solution of DEA + MDEA has been chosen as topic of the present study. In the work of Saghafi et al.[16], the RF model was developed and compared with modified Deshmukh–Mather method.[2] In the present effort, two new approaches including SGB trees and GP methodologies were employed to develop two new robust models over experimental data of Benamor and Aroua[2] and Haji-Sulaiman et al.[24] and were also compared with modified Deshmukh–Mather method[2] and the RF model.[16]

The main target of the present study was to produce accurate and acceptable models based on new approaches along with making comparison between different machine learning techniques.

Materials and methods

Dataset preparation

It is very important to apply a reliable dataset in order to develop an accurate and universal model. An experimental dataset including 175 sets of data of concentration of DEA (C_{DEA}), concentration of MDEA (C_{MDEA}), temperature (T), CO_2 partial pressure (P_{CO2}) and CO_2 loading capacity (α) in DEA, MDEA and their mixtures were collected from the study of Benamor and Aroua[2] and Haji-Sulaiman et al.[24] Table 1 shows the ranges of used experimental database for both independent and dependent variables.

A first preprocessing step is to divide the database into two subsets including training and test sets. The training set is used to evolve the model and the test proportion is employed to evaluate and test the prediction performance of new developed model. In this study, 85% of all data were randomly selected as training set and the remnant data were considered as independent test set. The distribution of dataset within these two phases was acceptable due to relatively uniform dispersion of the data on the domain of training as well as test set.

The next step is to feed the collected experimental data (*i.e.* the training and test datasets) to two robust machine learning algorithms in order to create new models. The considered algorithms in the present study are SGB trees and GP which are explained in the next sections.

Table 1. The ranges of used independent and dependent variables in this study.

Parameters	Range
C_{DEA} (mol/L)	0–4
C_{MDEA} (mol/L)	0–4
T (K)	303–323
P_{CO2} (KPa)	0.1–107.1
CO_2 loading α (mol/mol)	0.01–0.88

Stochastic gradient boosting (SGB) trees

In machine learning, single weak learners cannot produce a reliable model, but a combination of weak learners might well do. This is the fundamental idea behind gradient boosting (GB) framework,[25] a generic algorithm for producing a very accurate prediction rule by combining rough and moderately inaccurate rules-of-thumb.[26] Hence, GB can be considered as a series of ensemble learners consisting of many weak learners. A weak learner is defined as one whose performance is at least slightly better than random chance and decision trees (mostly regression trees) are used as the weak learners in GB framework.

Trees are constructed by choosing the best split points based on purity scores like Gini or based on minimizing the loss. Each tree focuses on the examples that the previous ones found difficult to classify.[27] The ultimate aim is to minimize the loss of the model using a gradient descent procedure. This generalized framework allows this technique to be applicable not just to the binary classification problems but also to regressions, multi-class classification, and many more.

Having said that the GB framework suffers from overfitting that might arise as a result of correlation between the trees (*i.e.* the weak learners). SGB addresses this problem by introducing randomization in the sampling of the training data.[28,29]

Proposed by J.H. Friedman,[30] SGB is inspired by the bagging technique[30,31] in the case of RF that is to allow the trees to be created from "subsamples" of the training dataset instead of total data. In other words, for each tree construction, a subsample of the training data is randomly drawn from the full dataset (without replacement) and the tree is then trained only on the subsample. This reduces the correlation between the trees, allows for self-validation of the model internally by using out of bag error estimates, and increases the accuracy and execution speed of the GB.[28,31] The literature has demonstrated the success of this new approach in various sectors and applications especially in chemical engineering areas.[32,33]

Performance of the SGB model is directly influenced by the optimal number of trees and learning rate. The specific weight with which sequential simple trees are added is referred to as the learning rate. So, best performance of the SGB model depends upon the optimum combination of these parameters. In the present study, the SGB model chose the optimum number of

trees as 1894 where the minimum error of test data (RMSD = 0.02298 mol/mol) was obtained (See Fig. 1). Also, 0.15 was selected as the optimum value for learning rate based on trial and error process.

Genetic programming (GP) methodology

GP as an effective soft computing approach for optimization and modeling purposes was introduced in the 1990s by John Koza.[34] GP algorithm has been patterned similar to biological generation mechanism which evolves evolutionary computer programs in its machine learning methodology to perform the tasks. In GP procedure, a population of mathematical functions is randomly generated from user-defined mathematical operators and some of the functions are arranged in the form of one or several "gene(s)" represented as chromosomes like syntactic tree structure(s) operating on input data. A simple schematic of gene is shown in Fig. 2. After generating the first population (also known as parents), the overall primary model is determined by weighted summation of all the genes with bias term.

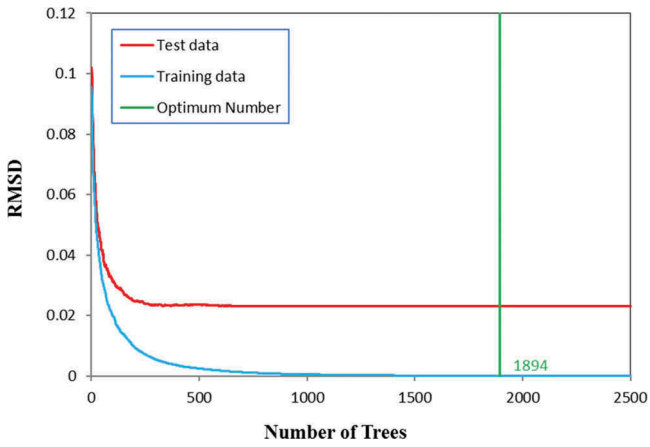

Figure 1. RMSD variation versus number of involved trees in the process of SGB algorithm. The optimum number of trees has been selected as 1894 at minimum RMSD of test dataset.

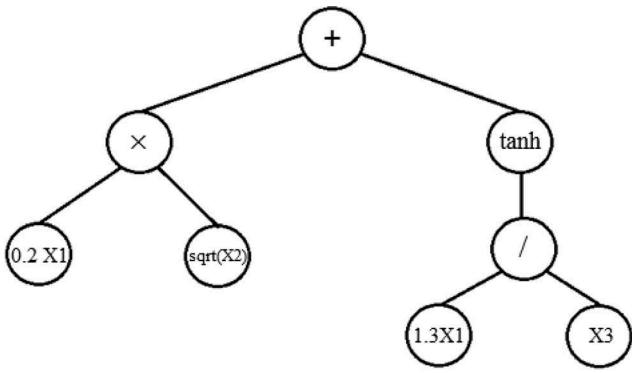

Figure 2. Simple schematic of GP tree structure (gene) including the operators: +, ×, tanh, sqrt, and /.

In the next step, the tree structures will be modified by crossing over the best operating trees and cutting some parts of trees and exchanging the cut parts. This modification is mainly performed to produce a new population (next generation or children). The generation procedure is iterated several times until the last population is generated which contains a number of functions in a specific arrangement solving the problem in the most-optimized manner.[7,35] The general flowchart of GP has been presented in Fig. 3.[36]

In the present study, number of first population and the number of generations were determined as 180 and 180, respectively and the mathematical operators employed in GP process were +, −, ×, /, and exp (*i.e.* natural logarithm).

Generally for modeling purposes, GP process performed to specify mathematical correlation between objective function and independent variables is known as "symbolic regression." In the case of symbolic regression using a number of genes rather than one, the algorithm is called "multi-gene symbolic regression" which is more applicable technique consisting of one or more genes (individual usual GP tree). Multi-gene method usually gives simpler functions with more flexible arrangement.[37–39]

Assessment of developed SGB and GP models

Different statistical criteria are available to evaluate the accuracy and applicability of a mathematical model. Root-mean-square deviation (RMSD), average absolute relative deviation percent (AARD %), and correlation coefficient (R^2) as well-known widely used parameters were applied in this study to assess the performance of the SGB and GP models. Lower values of RMSD and AARD (*i.e.* closer to zero) are preferred and the R^2 should be near to unity. These statistical parameters are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i^{\text{exp.}} - y_i^{\text{cal.}}\right)^2}{\sum_{i=1}^{n} \left(y_i^{\text{exp.}} - \overline{y}^{\text{exp.}}\right)^2} \tag{11}$$

$$\text{RMSD} = \sqrt{\left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(y_i^{\text{exp.}} - y_i^{\text{cal.}}\right)^2} \tag{12}$$

$$\text{ARD}(\%) = \left| \frac{\left(y_i^{\text{exp.}} - y_i^{\text{cal.}}\right)}{y_i^{\text{exp.}}} \right| \times 100 \tag{13}$$

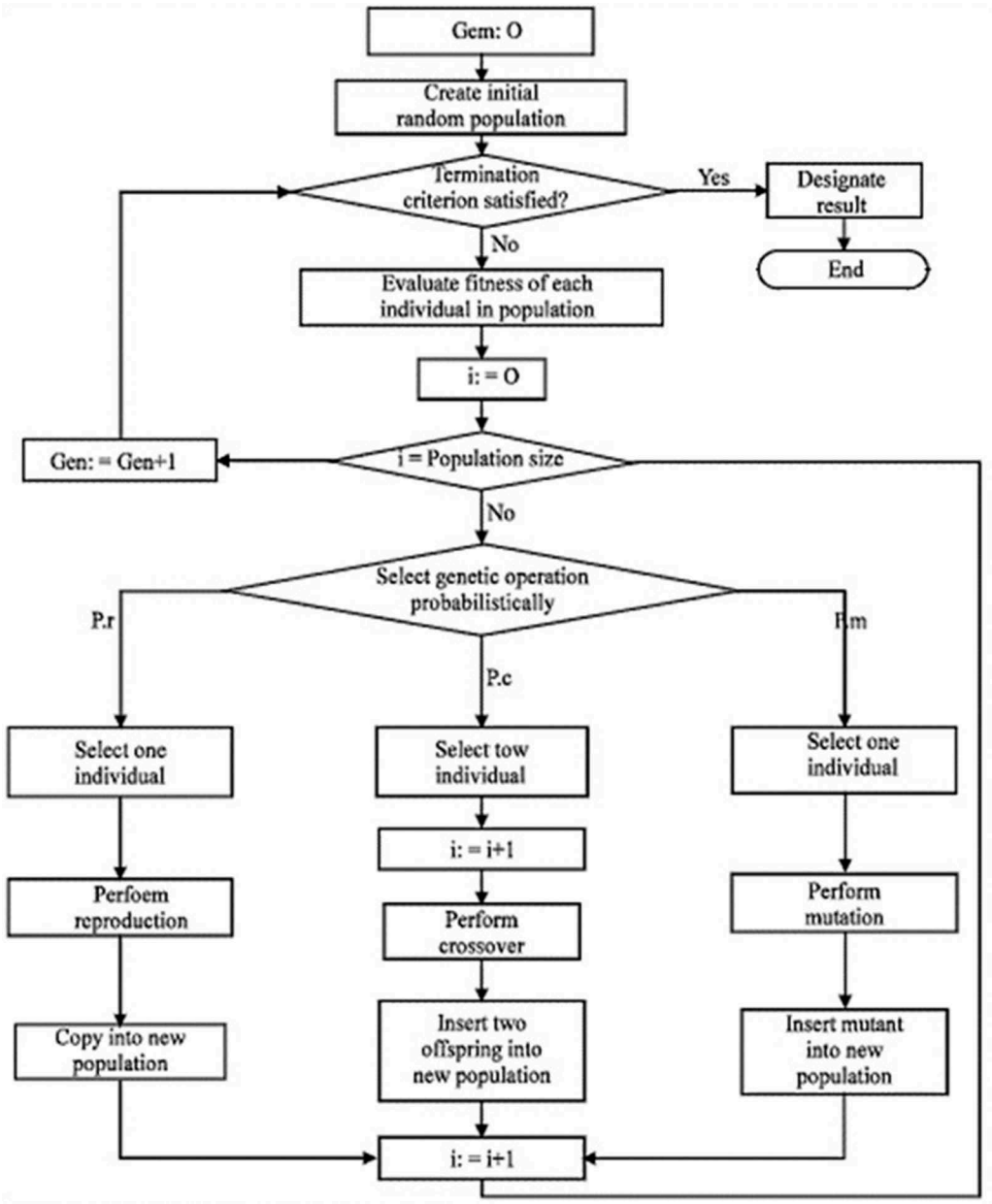

Figure 3. General flowchart of genetic algorithm.

$$\mathrm{AARD}(\%) = \left(\frac{1}{\mathrm{n}}\right) \sum_{\mathrm{i}=1}^{\mathrm{n}} \left| \frac{(y_i^{\mathrm{exp.}} - y_i^{\mathrm{cal.}})}{y_i^{\mathrm{exp.}}} \right| \times 100 \qquad (14)$$

$y_i^{exp.}$, $y_i^{cal.}$, $\overline{y}^{exp.}$ and n are experimental, estimated, average of experimental values of objective function (CO_2 loading capacity (α)), and number of samples in the dataset, respectively. The statistical parameters of the new SGB model and GP-based correlation for CO_2 loading capacity are presented in Table 2.

Results and discussion

By using SGB and GP methods, two new models were produced to estimate CO_2 loading capacity of amine (DEA, MDEA, and their mixtures) solution. The execution of the SGB algorithm in this study follows the explanations in Friedman[30,31] and the GP-based model is a mathematical correlation, as follows:

$$CO_2\text{Loading}\alpha = \left| \begin{array}{l} 0.004864(P_{CO_2}+\exp(C_{MDEA}-P_{CO_2}))+0.01376(C_{DEA}C_{MDEA}-C_{DEA}-\exp(-T-P_{CO_2})) \\ +0.07759(C_{DEA}\exp(-C_{DEA})-C_{MDEA})-0.829\exp(-P_{CO_2})+4.974\times10^{-6}\ T^2\exp(-P_{CO_2}) \\ -0.006191T+0.02482\exp(-C_{MDEA})+0.178874P_{CO_2}{}^2(\exp(-\exp(-P_{CO_2}))-1) \\ -0.0001813(C_{MDEA}+T)\exp(-C_{DEA})-4.298\times10^{-5}\ P_{CO_2}(C_{MDEA}+T+P_{CO_2})\exp(-C_{DEA}) \\ +0.02123P_{CO_2}\exp(-C_{DEA})+2.393 \end{array} \right| \quad (15)$$

Table 2. Statistical parameters of new developed models to determine the CO_2 loading capacity.

	SGB model			GP model		
Statistical parameters	All	Training	Test	All	Training	Test
AARD%	0.95628	0.01994	6.08881	8.71909	8.88598	7.76269
RMSD (mol/mol)	0.00903	0.00006	0.02298	0.02244	0.02169	0.02462
R^2	0.99848	1	0.99013	0.99087	0.99140	0.98739

Estimated values of CO_2 loading capacity in DEA, MDEA, and their mixtures along with the experimental data have been depicted in Figs. 4 and 5 for GP and SGB models, respectively. In these figures, all of the estimated data points have appropriately lied on the expected line (45-degree line). Table 2 represents the statistical parameters computed for all, training and test sets. The values of AARD%, RMSD, and R^2 have met the required standards and accuracy of interest for both GP and SGB methods. As another investigation, the absolute relative deviation (ARD) over all dataset have been drawn in the form of number of data points in different ranges of ARD % in Figs. 6 and 7 for GP and SGB models. In these figures, more number of data has lower errors. Based on Figs. 4–7 and Table 2, the results show acceptable accuracy and high estimation ability of the new developed models. The maximum ARD% of the SGB model is 25.1% and its absolute error is 0.0414 mol CO_2/mol amine. The maximum absolute error of SGB model in the dataset is 0.053 mol CO_2/mol amine with the ARD of 11.226%. The maximum ARD% for GP model is 104.37% which corresponds to absolute error of 0.028 mol CO_2 /mol amine. The maximum absolute deviation in the GP model is 0.072 mol CO_2/mol amine with the ARD of 30.25%. The results verify more applicability and accuracy of the SGB model.

The cumulative frequency percent (%) of the proposed models versus ARD% over all dataset has been shown in Fig. 8. In this graph, more applicability of the SGB model (blue curve) is clearly represented. 86.286% of all the CO_2 solubility data estimated by SGB approach have ARD% less than 1.272%, and 97.143% of all data have ARD% less than 6.775%. Indeed, there are just five data points with ARD% values larger than 10% in SGB model.

To evaluate the estimation capability of different methods, comparison of the new developed models along with RF model[16] and modified Deshmukh–Mather model[2]

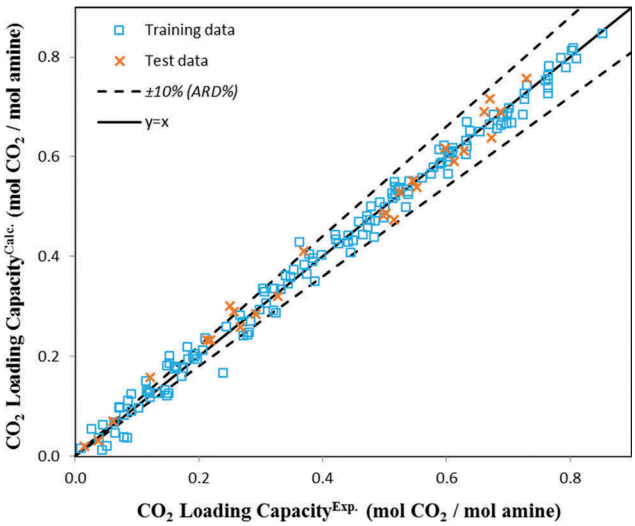

Figure 4. The scatterplot of GP model for both train and test partitions.

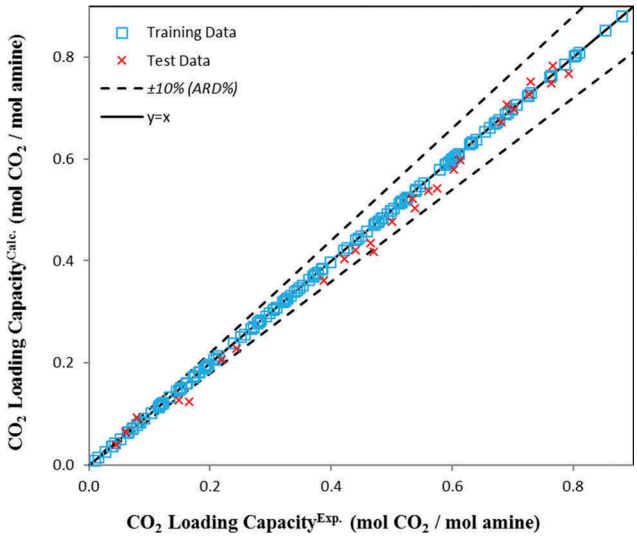

Figure 5. The scatterplot of SGB model for train and test datasets.

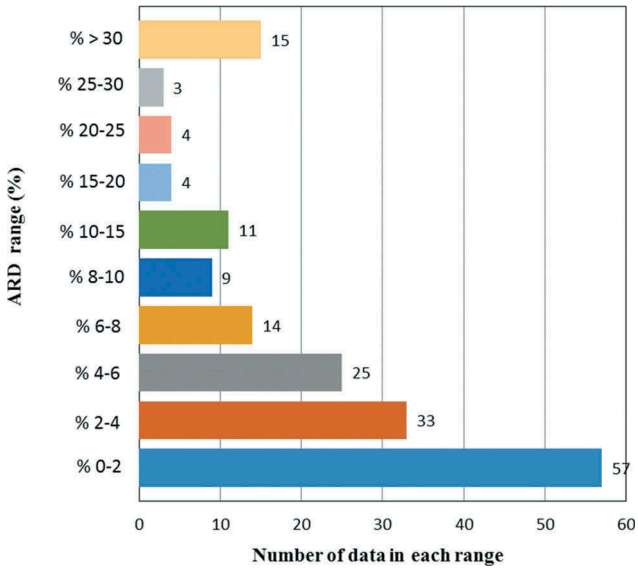

Figure 6. The absolute relative deviation percent (ARD%) of new GP model over total dataset. The maximum ARD% in all the dataset is 104.37% with absolute error of 0.028 [mol CO_2 /mol amine]. Maximum absolute error in dataset is 0.072 [mol CO_2/mol amine] with the ARD of 30.25%.

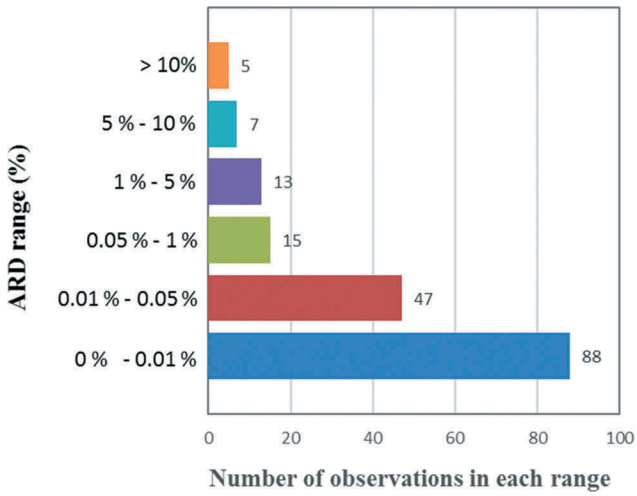

Figure 7. The absolute relative deviation percent (ARD%) of the SGB model of the present study over all dataset. The maximum ARD% in the dataset is 25.1% with absolute error of 0.0414 [mol CO_2/mol amine]. The maximum absolute error in the dataset is 0.053 [mol CO_2/mol amine] with the ARD of 11.226%.

has been implemented over same dataset (88 experimental data of RF model[16] and modified Deshmukh–Mather model[2]) and the result has been presented in Fig. 9. According to Fig. 9, SGB model provided the most accuracy with minimum AARD (0.94%) and Deshmukh–Mather model[2] presented the weakest estimations (AARD = 12.1%). Subsequently, it has been sufficiently demonstrated that the new SGB model is the most applicable method for confident prediction of CO_2 capturing in

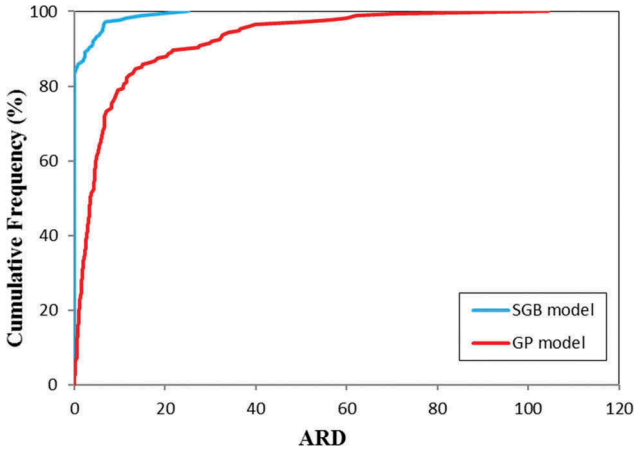

Figure 8. Cumulative frequency versus ARD% of the SGB and GP models over all dataset (175 data). 86.286% of all the estimated CO2 solubility by SGB have ARD% lower than 1.272% and 97.143% of all data have ARD% below 6.775%. There are only five data points with ARD% larger than 10% and the maximum ARD% over all dataset is 25.103% for SGB model.

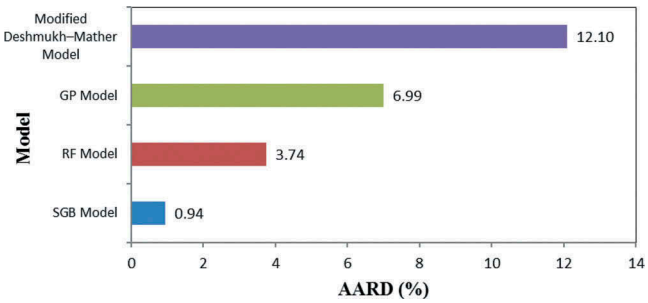

Figure 9. Comparison of different proposed models for estimating the CO_2 solubility in DEA + MDEA solution. The result of comparison demonstrated that the new developed SGB model provided the most accuracy with minimum AARD%.

amines of DEA, MDEA, and DEA + MDEA among different models.

More analysis of the new SGB and GP models has been given in Fig. 10 which presents experimental and estimated data of CO_2 loading capacity versus CO_2 partial pressure at different conditions. The plots (a), (b), and (c) have been drawn in fixed solution of (0.5M DEA + 1.5M MDEA), (1M DEA + 1M MDEA) and (1.5M DEA + 0.5M MDEA), respectively. In each plot, three curves have been considered for different temperatures of 303, 313, and 323 K. Figure 10 shows that CO_2 solubility increases with enhancing CO_2 partial pressure and decreases with enhancing the temperature. It has been demonstrated in Fig. 10 that the outputs of both SGB and GP models are well-matched with experimental data and have acceptable trends.

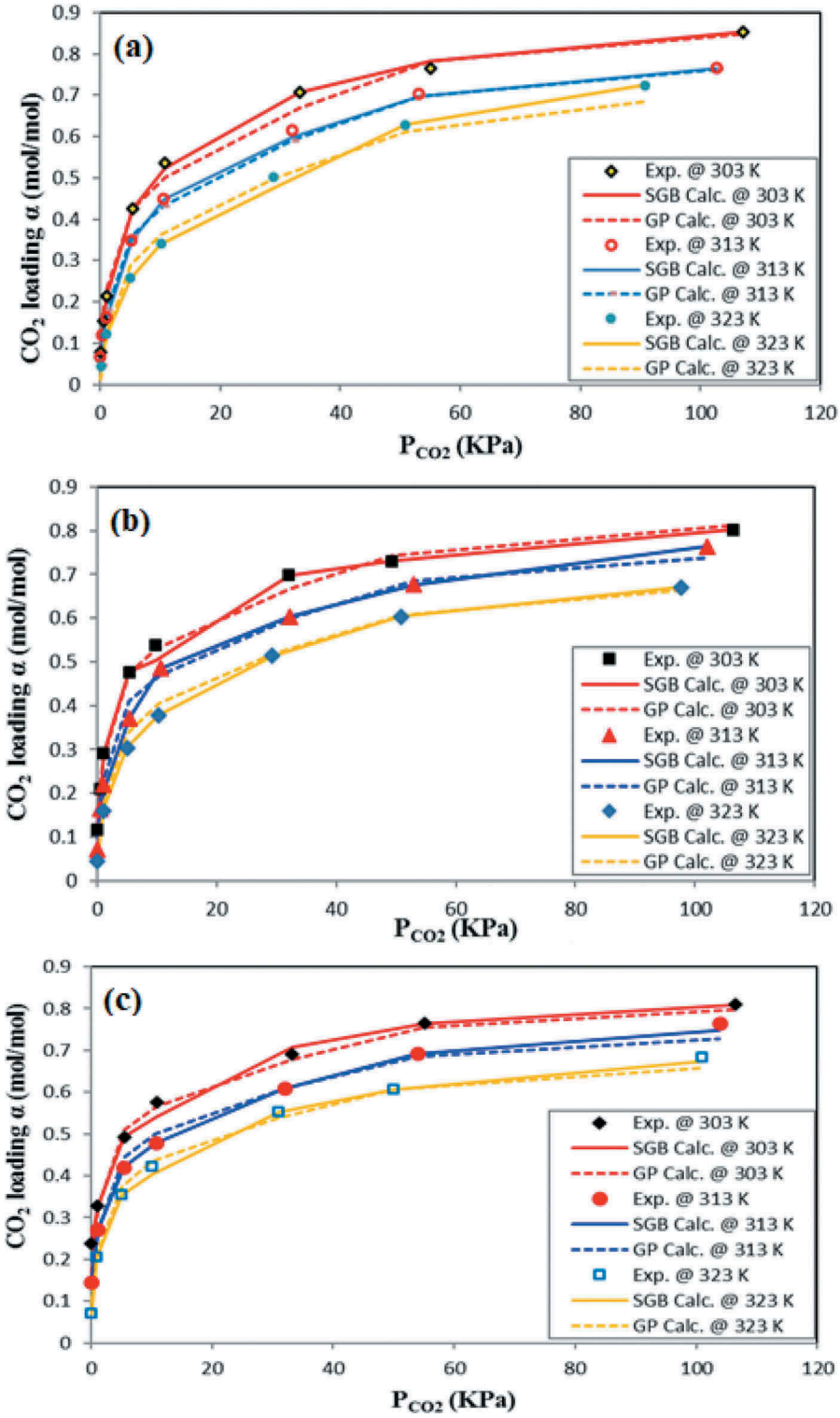

Figure 10. Trend of CO_2 loading capacity with CO_2 pressure in different solutions of (a) 0.5M DEA + 1.5M MDEA (b) 1M DEA + 1M MDEA, and (c) 1.5M DEA + 0.5M MDEA. The SGB-estimated data (curves) and GP-estimated data (dashed) are in good agreement with experimental data (points) and show acceptable trends of variables.

The relative importance of independent variables (C_{DEA}, C_{MDEA}, T, and P_{CO2}) with respect to the predictability of the objective function (CO_2 loading capacity) in the SGB model developing procedure was depicted in Fig. 11. A higher value of relative importance indicates the stronger influence in the model. It is observed in Fig. 11 that each variable contributes to the objective function in a different way and the partial

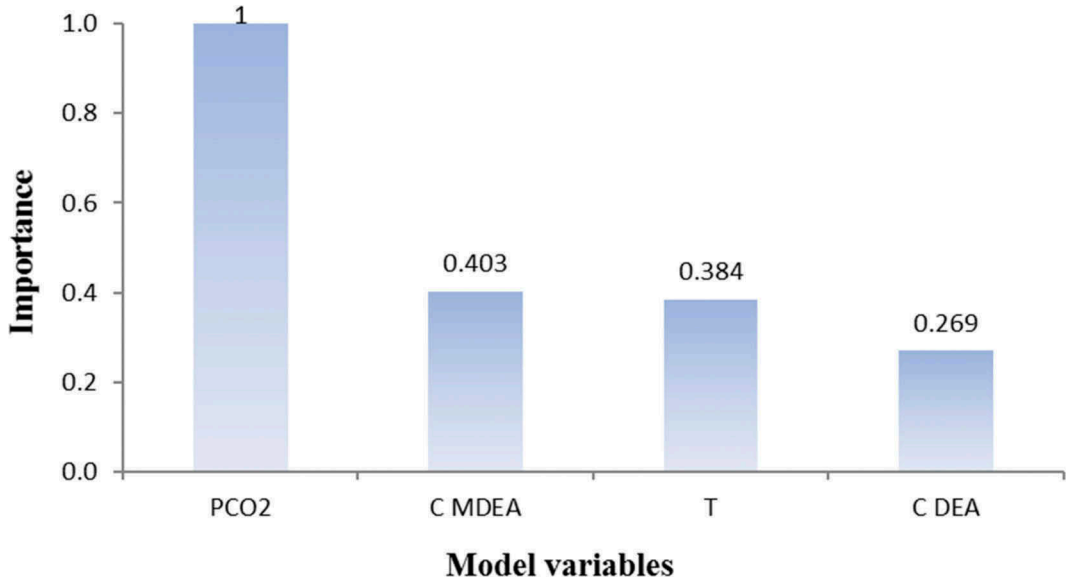

Figure 11. Relative importance of the independent variables in the CO_2 solubility model by SGB analysis.

pressure of CO_2 is the most influencing factor. This result is consistent with the outcome of RF model proposed by Saghafi et al.[16]

It is recommended using SGB model as a smart reliable tool by chemical engineers and everybody tackling the problem of CO_2 solubility prediction in DEA, MDEA, and DEA + MDEA solutions. This type of model could especially be applied in processing/sweetening units of natural gas as a great and highly used source of fossil energy and has important roles to minimize the CO_2 content in the gas.

Conclusion

In this inquiry, two smart machine learning algorithms, viz. SGB trees and GP, were used for estimation of CO_2 loading capacity in the solutions of DEA, MDEA, and DEA +MDEA based on the presented experimental data. The new proposed models correlate the CO_2 loading capacity with DEA concentration, MDEA concentration, temperature, and CO_2 partial pressure. Statistical assessment showed that machine learning frameworks used in the present study provide engaging and reliable methods that may engineers to fruitfully predict future trends in CO_2 solubility for development and design of new sour gas treatment system purposes. However, different analyses recommend that SGB scheme generate more accurate predictions than GP and other models previously published (*i.e.* RF and Modified Deshmukh–Mather approaches) for the same purpose. Based on the obtained results, it can be concluded that SGB model can give reasonable prediction performance and may be utilized to ascertain theses predicting approaches that could provide precise and dependable CO_2 solubility prediction in aqueous solutions of DEA and MDEA. The current paper is also the first to utilize SGB scheme in CO_2 solubility predicting. Finally, the relative importance analysis was conducted and it was found that the partial pressure of CO_2 was the most important one among all independent variables.

Nomenclature

AARD %	Average absolute relative deviation
AMP	2-amino, 2-methyl-1-propanol
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
ARD%	Absolute relative deviation
C_{DEA}	Concentration of DEA
C_{MDEA}	Concentration of MDEA
DEA	Di-ethanol amine
DIPA	Di-isopropanol amine
E	Extract phase mass (or mol) rate
G	Gas feed mass (or mol) rate
GB	Gradient boosting
GP	Genetic programming
GRN	Generalized regression neural networks
ICA	Imperialist competitive algorithm
LSSVM	Least squares support vector machine
MDEA	Methyl, di-ethanol amine
MEA	Mono-ethanol amine
M	Molar concentration (mol/L)
n	Number of samples in the dataset
P_{CO2}	Carbon dioxide gas partial pressure
PSO	Particle swarm optimization
PZ	Piperazine
R	Raffinate phase mass (or mol) rate
R^2	Squared correlation coefficient
RMSD	Root-mean-square deviation

S	Solvent phase mass (or mol) rate
SGB	Stochastic gradient boosting
SVM	Support vector machine
T	Temperature
TEA	Tri-ethanol amine
x_1	Mass (or mol) fraction of CO_2 in solvent
x_2	Mass (or mol) fraction of CO_2 in raffinate phase
y_1	Mass (or mol) fraction of CO_2 in gas feed
y_2	Mass (or mol) fraction of CO_2 in extract phase
$y_i^{cal.}$	Predicted dependent variable
$y_i^{exp.}$	Experimental dependent variable
$\overline{y}^{exp.}$	Average of experimental dependent variable
α	CO_2 loading capacity

References

[1] Maddox, R.N.; (1974) *Gas and liquid sweetening*;JM Campbell for Campbell Petroleum series

[2] Benamor, A.; Aroua, M. (2005) Modeling of CO 2 solubility and carbamate concentration in DEA, MDEA and their mixtures using the Deshmukh–Mather model. *Fluid Phase Equilibria*, 231 (2): 150–162. doi:10.1016/j.fluid.2005.02.005

[3] Treybal, R.E. (1980) *Mass Transfer Operations*, McGraw-Hill Book Co, Singapore.

[4] Hsu, C.H.; Chu, H.; Cho, C.M. (2003) Absorption and reaction kinetics of amines and ammonia solutions with carbon dioxide in flue gas. *Journal of the Air & Waste Management Association*, 53 (2): 246–252. doi:10.1080/10473289.2003.10466139

[5] Bougie, F.; Iliuta, M.C. (2012) Sterically hindered amine-based absorbents for the removal of CO2 from gas streams. *Journal of Chemical & Engineering Data*, 57 (3): 635–669. doi:10.1021/je200731v

[6] Porcheron, F.; Gibert, A.; Mougin, P.; Wender, A.L. (2011) High throughput screening of CO2 solubility in aqueous monoamine solutions. *Environmental Science & Technology*, 45 (6): 2486–2492. doi:10.1021/es103453f

[7] Abooali, D.; Khamehchi, E. (2017) New predictive method for estimation of natural gas hydrate formation temperature using genetic programming. *Neural Computing and Applications*, https://doi.org/10.1007/s00521-017-3208-0.

[8] Abooali, D.; Khamehchi, E. (2014) Estimation of dynamic viscosity of natural gas based on genetic programming methodology. *Journal of Natural Gas Science and Engineering*, 21: 1025–1031. doi:10.1016/j.jngse.2014.11.006

[9] Ahmadi, M.; Soleimani, R.; Bahadori, A. (2016) Prediction of natural gas hydrate inhibitor vaporization rate using particle swarm optimization approach. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 38 (12): 1706–1712. doi:10.1080/15567036.2014.975298

[10] Ahmadi, M.A.; Haghbakhsh, R.; Soleimani, R.; Bajestani, M.B. (2014) Estimation of H 2 S solubility in ionic liquids using a rigorous method. *The Journal of Supercritical Fluids*, 92: 60–69. doi:10.1016/j.supflu.2014.05.003

[11] Ahmadi, M.A.; Soleimani, R.; Bahadori, A. (2014) A computational intelligence scheme for prediction equilibrium water dew point of natural gas in TEG dehydration systems. *Fuel*, 137: 145–154. doi:10.1016/j.fuel.2014.07.072

[12] Ahmadi, M.A.; Soleimani, R.; Lee, M.; Kashiwao, T.; Bahadori, A. (2015) Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool. *Petroleum*, 1 (2): 118–132. doi:10.1016/j.petlm.2015.06.004

[13] Ahmadi, M.-A.; Pouladi, B.; Javvi, Y.; Alfkhani, S.; Soleimani, R. (2015) Connectionist technique estimates H 2 S solubility in ionic liquids through a low parameter approach. *The Journal of Supercritical Fluids*, 97: 81–87. doi:10.1016/j.supflu.2014.11.009

[14] Ghiasi, M.M.; Mohammadi, A.H. (2014) Rigorous modeling of CO 2 equilibrium absorption in MEA, DEA, and TEA aqueous solutions. *Journal of Natural Gas Science and Engineering*, 18: 39–46. doi:10.1016/j.jngse.2014.01.005

[15] Hashemkhani, M.; Soleimani, R.; Fazeli, H.; Lee, M.; Bahadori, A.; Tavalaeian, M. (2015) Prediction of the binary surface tension of mixtures containing ionic liquids using support vector machine algorithms. *Journal of Molecular Liquids*, 211: 534–552. doi:10.1016/j.molliq.2015.07.038

[16] Saghafi, H.; Ghiasi, M.M.; Mohammadi, A.H. (2017) Analyzing the experimental data of CO2 equilibrium absorption in the aqueous solution of DEA+ MDEA with Random Forest and Leverage method. *International Journal of Greenhouse Gas Control*, 63: 329–337. doi:10.1016/j.ijggc.2017.03.028

[17] Shafiei, A.; Ahmadi, M.A.; Zaheri, S.H.; Baghban, A.; Amirfakhrian, A.; Soleimani, R. (2014) Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach. *The Journal of Supercritical Fluids*, 95: 525–534. doi:10.1016/j.supflu.2014.08.011

[18] Daneshvar, N.; Moattar, M.Z.; Abdi, M.A.; Aber, S. (2004) Carbon dioxide equilibrium absorption in the multi-component systems of CO 2+ TIPA+ MEA+ H 2 O, CO 2+ TIPA+ Pz+ H 2 O and CO 2+ TIPA+ H 2 O at low CO 2 partial pressures: experimental solubility data, corrosion study and modeling with artificial neural network. *Separation and Purification Technology*, 37 (2): 135–147.

[19] Sipöcz, N.; Tobiesen, F.A.; Assadi, M. (2011) The use of artificial neural network models for CO 2 capture plants. *Applied Energy*, 88 (7): 2368–2376. doi:10.1016/j.apenergy.2011.01.013

[20] Shahsavand, A.; Fard, F.D.; Sotoudeh, F. (2011) Application of artificial neural networks for simulation of experimental CO 2 absorption data in a packed column. *Journal of Natural Gas Science and Engineering*, 3 (3): 518–529. doi:10.1016/j.jngse.2011.05.001

[21] Pahlavanzadeh, H.; Nourani, S.; Saber, M. (2011) Experimental analysis and modeling of CO 2 solubility in AMP (2-amino-2-methyl-1-propanol) at low CO 2 partial pressure using the models of Deshmukh–Mather and the artificial neural network. *The Journal of Chemical Thermodynamics*, 43 (12): 1775–1783. doi:10.1016/j.jct.2011.05.032

[22] Zhou, Q.; Chan, C.; Tontiwachwuthikul, P.; Idem, R.; Gelowitz, D. (2013) Application of neuro-fuzzy modeling technique for operational problem solving in a CO 2 capture process system. *International Journal of*

Greenhouse Gas Control, 15: 32–41. doi:10.1016/j.ijggc.2013.01.031

[23] Zhou, Q.; Wu, Y.; Chan, C.W.; Tontiwachwuthikul, P. (2011) From neural network to neuro-fuzzy modeling: applications to the carbon dioxide capture process. *Energy Procedia*, 4: 2066–2073. doi:10.1016/j.egypro.2011.02.089

[24] Haji-Sulaiman, M.; Aroua, M.; Benamor, A. (1998) Analysis of equilibrium data of CO2 in aqueous solutions of diethanolamine (DEA), methyldiethanolamine (MDEA) and their mixtures using the modified Kent Eisenberg model. *Chemical Engineering Research and Design*, 76 (8): 961–968. doi:10.1205/026387698525603

[25] Kearns, M. (1988) Thoughts on hypothesis boosting. *Unpublished manuscript*, 45: 105.

[26] Freund, Y.; Schapire, R.E.(1995) A desicion-theoretic generalization of on-line learning and an application to boosting. In: Vitányi P. (eds) *Computational Learning Theory. EuroCOLT 1995. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol 904. Springer, Berlin, Heidelberg

[27] Valiant, L.(2013) *Probably Approximately Correct: Natureõs Algorithms for Learning and Prospering in a Complex World*. New York, NY*: Basic Books*.

[28] Kriegler, B.; Berk, R. (2010) Small area estimation of the homeless in Los Angeles: an application of cost-sensitive stochastic gradient boosting. *The Annals of Applied Statistics*, 1234–1255. doi:10.1214/10-AOAS328

[29] Kuhn, M.; Johnson, K.(2013) *Applied Predictive Modeling*, New York, NY: Springer.

[30] Friedman, J.H. (2002) Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38 (4): 367–378. doi:10.1016/S0167-9473(01)00065-2

[31] Friedman, J.H. (2001) Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 1189–1232. doi:10.1214/aos/1013203451

[32] Soleimani, R.; Dehaghani, A.H.S.; Bahadori, A. (2017) A new decision tree based algorithm for prediction of hydrogen sulfide solubility in various ionic liquids. *Journal of Molecular Liquids*, 242: 701–713. doi:10.1016/j.molliq.2017.07.075

[33] Soleimani, R.; Mahmood, T.; Bahadori, A. (2016) Assessment of compressor power and condenser duty per refrigeration duty in three-stage propane refrigerant systems using a new ensemble learning tool. *Chemeca 2016: Chemical Engineering-Regeneration, Recovery and Reinvention*, 23, Adelaide SA, Australia.

[34] Koza, J. (1992) *Genetic Programming*,Massachusetts Institute of Technology. In MIT Press: London, UK.

[35] Abooali, D.; Khamehchi, E. (2016) Toward predictive models for estimation of bubble-point pressure and formation volume factor of crude oil using an intelligent approach. *Brazilian Journal of Chemical Engineering*, 33 (4): 1083–1090. doi:10.1590/0104-6632.20160334s20150374

[36] Soleimani, R.; Abooali, D.; Shoushtari, N.A. (2018) Characterizing CO2 capture with aqueous solutions of LysK and the mixture of MAPA+ DEEA using soft computing methods. *Energy*, 164: 664–675. doi:10.1016/j.energy.2018.09.061

[37] Abooali, D.; Sobati, M.A.; Shahhosseini, S.; Assareh, M. (2019) A new empirical model for estimation of crude oil/brine interfacial tension using genetic programming approach. *Journal of Petroleum Science and Engineering*, 173: 187–196. doi:10.1016/j.petrol.2018.09.073

[38] Searson, D.; (2009) GPTIPS: genetic programming & symbolic regression for MATLAB. *User Guide*, 2010.

[39] Searson, D.P.; Leahy, D.E.; Willis, M.J. (2010) GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. In: *Proceedings of the International Multiconference of Engineers and Computer Scientists*, Vol. 1; 77–80, Hong Kong, IMECS.