

# Applying ANN, ANFIS, and LSSVM Models for Estimation of Acid Solvent Solubility in Supercritical CO<sub>2</sub>

Amin Bemani<sup>1</sup>, Alireza Baghban<sup>2</sup>, Shahaboddin Shamshirband<sup>3,4</sup>, Amir Mosavi<sup>5,6</sup>, Peter Csiba<sup>7</sup>, Annamária R. Várkonyi-Kóczy<sup>5,7</sup>

<sup>1</sup> Petroleum Engineering Department, Petroleum University of Technology, Ahwaz, Iran; [aminbemani90@yahoo.com](mailto:aminbemani90@yahoo.com)

<sup>2</sup> Chemical Engineering Department, Amirkabir University of Technology, Mahshahr Campus, Mahshahr, Iran; [baghban1369@gmail.com](mailto:baghban1369@gmail.com)

<sup>3</sup> Department for Management of Science and Technology Development, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

<sup>4</sup> Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

<sup>5</sup> Department of Automation, Óbuda University, Budapest, Hungary

<sup>6</sup> School of the Built Environment, Oxford Brookes University, Oxford OX3 0BP, UK  
[amir.mosavi@kvk.uni-obuda.hu](mailto:amir.mosavi@kvk.uni-obuda.hu)

<sup>7</sup> Department of Mathematics and Informatics, J. Selye University, Komarno 94501, Slovakia  
[csibap@uj.s.sk](mailto:csibap@uj.s.sk) and [koczya@uj.s.sk](mailto:koczya@uj.s.sk)

## Abstract

In the present work, a novel and the robust computational investigation is carried out to estimate solubility of different acids in supercritical carbon dioxide. Four different algorithms such as radial basis function artificial neural network, Multi-layer Perceptron (MLP) artificial neural network (ANN), Least squares support vector machine (LSSVM) and adaptive neuro-fuzzy inference system (ANFIS) are developed to predict the solubility of different acids in carbon dioxide based on the temperature, pressure, hydrogen number, carbon number, molecular weight, and acid dissociation constant of acid. In the purpose of best evaluation of proposed models, different graphical and statistical analyses and also a novel sensitivity analysis are carried out. The present study proposed the great manners for best acid solubility estimation in supercritical carbon dioxide, which can be helpful for engineers and chemists to predict operational conditions in industries.

**Keywords:** Supercritical carbon dioxide, machine learning modeling, acid, artificial intelligence, solubility, artificial neural networks (ANN), adaptive neuro-fuzzy inference system (ANFIS), Least-squares support-vector machine (LSSVM), Multi-layer Perceptron (MLP), engineering applications of artificial intelligence

## 1 Introduction

In the recent years, supercritical fluid has become one of the interests of chemical engineers and chemists as a novel and extensive applicable technology. The synthesis and generating of nanomaterials and extraction process of different materials are the popular applications of supercritical fluids (Inomata, Honma et al. 1999, Stassi, Bettini et al. 2000, Ohde, Hunt et al. 2001, Celso, Triolo et al. 2002, Üzer, Akman et al. 2006, Munshi and Bhaduri 2009, Nahar and Sarker 2012, Zhang, Heinonen et al. 2014, Knez, Cör et al. 2017, Zhao, Zhang et al. 2017, Belghait, Si-Moussa et al. 2018, Daryasafar, Daryasafar et al. 2018, Gao, Abdi-khanghah et al. 2018). One of the supercritical fluids which have wide applications in the extraction of various metals from solid and liquid phases is carbon dioxide (Erkey 2000, Sunarso and Ismadji 2009, Lin, Liu et al. 2014). Due to non-flammability, nontoxicity, low cost, and critical points (304.2 K and 7.38 MPa) of carbon dioxide, the supercritical carbon dioxide becomes one of the interesting and applicable supercritical fluids in industries (Ghaziaskar and Nikravesh 2003, Bovard, Abdi et al. 2017). The viscosity and density of supercritical carbon dioxide are known as two important transport properties of the fluids which are affected by pressure and temperature. Another dominant thermos physical property of supercritical carbon dioxide is solubility of different materials in supercritical carbon dioxide which is a function of various factors such as polarity, molecular weight, pressure, temperature, and vapor pressure (Huang, Chiew et al. 2005, Ghaziaskar and Kaboudvand 2008). One types of the materials which have a solubility in supercritical carbon dioxide are acids, the nanofluoropentanoic acid which is known as one type of perfluorocarboxylic acids, has extensive applications in the production of paints additives, polymers, foams, and stain repellents but because of their high ability instability they are harmful to environment (Richter and Dibble 1983, Moody and Field 1999, Hintzer, Löhr et al. 2004, Fei and Olsen 2011, Hubbard, Guo et al. 2012, Dartiguelongue, Leybros et al. 2016, Hintzer, Juergens et al. 2016). Adrien Dartiguelongue and coworkers studied solubility of perfluoropentanoic acid in supercritical carbon dioxide in the wide range of temperature and pressure and also proposed some density based models to predict solubility in terms of density of supercritical fluids (Dartiguelongue, Leybros et al. 2016). Gurdial et al. constructed dynamic setup to study solubility of o-, m- and p-hydroxybenzoic acid in the supercritical carbon dioxide in the wide range pressure of 80-205 mbar and temperature range of 308.15-328.15 K and correlated the measured solubility as a function of density (Gurdial and Foster 1991). Kumoro measured the solubility of 2R,3 $\beta$ -dihydroxyurs-12-en-28-oic acid which is called Corosolic acid dynamically in a different range of pressure 8 to 30 MPa and five different temperatures of 308.15, 313.15, 323.15, and 333.15 K. Kumoro used various density based models to correlate the experimental data (Kumoro 2011).

Sahihi et al. measured the solubility of Maleic acid in supercritical carbon dioxide by utilization of static experimental setup. The measured data belongs to Maleic acid in pressure range of 7 to 300 bar and temperature of 348.15 K (Sahihi, Ghaziaskar et al. 2010). Ghaziaskar and coworkers used a continuous flow set up to study solubility of tracetin, diacentin and acetic acid in supercritical carbon dioxide in the pressure range of 70 to 180 bar and various temperature of 313, 333 and 348 K and they also compared the obtained solubilities for different acids (Ghaziaskar, Afsari et al. 2017). Helena Sovova adjusted the Adachi-Lu equation based on the solubility of Ribes nigrum (blackcurrant) and Vitis vinifera (grape-vine) in supercritical carbon dioxide. They concluded the Adachi-Lu equation has enough accuracy in forecasting solubility of triglycerides in carbon dioxide (Sovova, Zarevucka et al. 2001).

The issue of prediction of various acids solubility in supercritical carbon dioxide and phase equilibrium investigation of supercritical carbon dioxide and different materials are the important topics in chemical engineering research. According to the hardships of experimental studies such as special tools and procedure which are needed, in the present work, the mathematical investigation is considered as a great solution for these problems (Anitescu, Atroshchenko et al. 2019, Guo, Zhuang et al. 2019, Rabczuk, Ren et al. 2019, Zarei, Razavi et al. 2019). In this paper four different algorithms, Radial basis function artificial neural network (RBF-ANN), Multi-layer Perceptron artificial neural network (MLP-ANN), Least squares support vector machine (LSSVM) and Adaptive neuro-fuzzy inference system (ANFIS) are developed to predict the solubility of different types of acid in supercritical carbon dioxide based on the various parameters such as structure of acid, pressure and temperature.

## **2 Methodology**

### ***2.1 Experimental Data Gathering***

The dominant purpose of present paper is development of accurate and simple models to forecast solubility of different acids in supercritical carbon dioxide. Due to this, the required actual data for training and testing phases of models were assembled from the reliable source existed in literature (Gurdial and Foster 1991, Sovova, Zarevucka et al. 2001, Sparks, Hernandez et al. 2007, Tian, Jin et al. 2007, Sparks, Estévez et al. 2008, Kumoro 2011, Dartiguelongue, Leybros et al. 2016). This collection of data contains the 188 acid solubility data points in terms of pressure, temperature, acid dissociation constant, molecular weight, number of carbon and hydrogen of acid. The details of data collection are reported in Tab. S1 and Tab. S2. This details include acid name, acid dissociation constant, pressure and temperature ranges and number of utilized data points for each acid. Also, for clarification of this experimental dataset, the structure, linear formula and

molecular weight of utilized acids are presented in **Tab. S3**. These acids include Perfluoropentanoic acid, o-Hydroxybenzoic Acid, Corosolic Acid, Maleic Acid, Ferulic Acid, Azelaic Acid, p-aminobanzoic acid and Nonanioc acid.

## ***2.2 Artificial neural networks***

Artificial neural networks have amazing similarities to the performance and structure of neuron units in the brain system (Smith 1993, Baş and Boyacı 2007). These computational blocks construct different types of layer such as input, output and hidden layers. In the layers, there are transfer functions or activation function which organize the process of training in the algorithm. Each neuron has specific weight and bias values which control the optimization process. Artificial neural network has ability of tracing a nonlinear form relationship between input and output parameters. Due to this ability, artificial neural networks have widespread application in different industries and sciences.

Artificial neural networks can be classified in different forms such as a recurrent neural network (RNN), radial basis function and multilayer perceptron (Movagharnejad, Mehdizadeh et al. 2011, Abdi-Khanghah, Bemani et al. 2018, Zamen, Baghban et al. 2019). In the present work, the MLP and RBF network are utilized.

## ***2.3 Least squares support vector machine***

Vapnik organized support vector machine based on statistical learning theory (Vapnik 1998). This computational intelligence can be used for regression and classification purposes. However, there are many advantages to this method but there is a hardship in its computational procedure because of quadratic programming. The least squares SVM (LSSVM) is proposed as a novel type of SVM to solve this problem. This novel approach organized linear equations for computation and optimization (Cortes and Vapnik 1995, Suykens and Vandewalle 1999, Suykens, Vandewalle et al. 2001, Zamen, Baghban et al. 2019).

By considering a dataset of  $(x_i, y_i)_n$ , the LSSVM regression prediction is utilized to estimate a function, where  $x_i$  and  $y_i$  are known as input and target parameters and  $n$  represent the number of data which utilized in training phase (Wang, Zhang et al. 2005). The linear regression is formulated such as following:

$$y = \omega^T \varphi(x) + b \tag{1} \text{Eq.}$$

Where  $\varphi(x)$  denotes a nonlinear function that has different forms such as polynomial, linear, sigmoid and radial basis functions. Also,  $\omega$  and  $b$  denote the weights and determined constant

coefficient in training process. A new optimization problem can be defined based on LSSVM approach (Baghban, Bahadori et al. 2016, Baghban, Namvarrechi et al. 2016, Ahmadi, Baghban et al. 2019):

$$\frac{\min}{\omega, b, e} J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{k=1}^N e_k^2 \quad \text{Eq. (2)}$$

Which is related to the below constraints:

$$y_k = \omega^T \varphi(x_k) + b + e_k \quad k=1,2,\dots,N \quad \text{Eq. (3)}$$

The Lagrangian equation is constructed to solve the optimization problem:

$$L(\omega, b, e, \alpha) = J(\omega, e) - \sum_{k=1}^N \alpha_k \{\omega^T \varphi(x_k) + b + e_k - y_k\} \quad \text{Eq. (4)}$$

Where  $\gamma$  and  $e_k$  are known as regularization parameter and regression error. The  $\alpha_k$  represent the support value. To solve the above problem, the above equation is differentiated with respect to the different parameters:

$$\frac{\partial L(\omega, b, e, \alpha)}{\partial \omega} = 0 \rightarrow \omega = \sum_{k=1}^N \alpha_k \varphi(x_k) \quad \text{Eq. (5)}$$

$$\frac{\partial L(\omega, b, e, \alpha)}{\partial b} = 0 \rightarrow \sum_{k=1}^N \alpha_k = 0 \quad \text{Eq. (6)}$$

$$\frac{\partial L(\omega, b, e, \alpha)}{\partial e_k} = 0 \rightarrow \alpha_k = \gamma e_k, \quad k=1,2,\dots,N \quad \text{Eq. (7)}$$

$$\frac{\partial L(\omega, b, e, \alpha)}{\partial \alpha_k} = 0 \rightarrow y_k = \omega^T \varphi(x_k) + b + e_k \quad k=1,2,\dots,N \quad \text{Eq. (8)}$$

Karush–Kuhn–Trucker matrix can be obtained by elimination of  $\omega$  and  $e$  (Cortes and Vapnik 1995, Baylar, Hanbay et al. 2009, Mehdizadeh and Movagharnejad 2011):

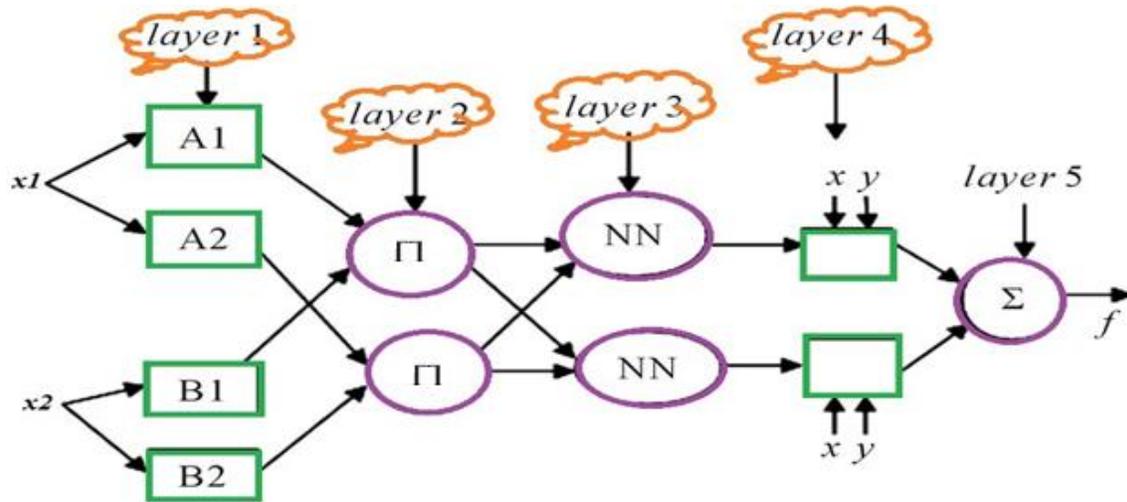
$$\begin{bmatrix} 0 & 1_v^T \\ 1_v & \Omega + \gamma^{-1}I \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad \text{Eq. (9)}$$

where  $y = [y_1 \dots y_N]^T$ ,  $\alpha = [\alpha_1 \dots \alpha_N]^T$ ,  $1_N = [1 \dots 1]^T$ , and  $I$  represents the identity matrix.  $\Omega_{kl}$  is  $\varphi(x_k)^T \varphi(x_l) = K(x_k, x_l)$ .  $K(x_k, x_l)$  is known as kernel function which can be in different forms of linear, polynomial and radial basis function forms (Gunn 1998). The estimating function form of LSSVM algorithm can be expressed as following formulation (Muller, Mika et al. 2001, Rostami, Baghban et al. 2019):

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad \text{Eq. (10)}$$

### 2.4 Adaptive neuro-fuzzy inference system (ANFIS)

Adaptive neuro-fuzzy inference system which is called ANFIS algorithm, in brief, has five different layers. The aforementioned approach was developed by Jang and Sun (Jang, Sun et al. 1997). The hybrid learning approach and back propagation are known as fundamentals of training of conventional ANFIS algorithm. The ANFIS algorithm was born base on fuzzy logic and neural network advantages and also the different evolutionary methods such as Imperialist Competitive Algorithm (ICA), Particle Swarm Optimization (PSO) and Genetic algorithm (GA) can be used to reach the optimal structure of ANFIS algorithm (Afshar, Gholami et al. 2014, Khosravi, Nunes et al. 2018, Razavi, Sabaghmoghadam et al. 2019). The ANFIS structure is demonstrated in **Fig. 1**. As shown there are two input variables and one output.



**Figure 1:** Typical construction of ANFIS approach

In the first layer, the linguistic terms are built based on input data. The Gaussian membership function is applied to organize these linguistic terms. The Gaussian function can be shown as following formulation (Ahangari, Moeinossadat et al. 2015, Bahadori, Baghban et al. 2016):

$$O_i^1 = \beta(X) = \exp\left(-\frac{1}{2} \frac{(X-Z)^2}{\sigma^2}\right) \quad \text{Eq. (11)}$$

Where  $Z$  and  $\sigma$  denote the Gaussian parameters.

The next layer, shown as  $\Pi$  multiplies the incoming signals and contains the weighted terms which are related to rules:

$$O_i^2 = W_i = \beta_{A_i}(X) \cdot \beta_{B_i}(X) \quad \text{Eq. (12)}$$

The third layer the shown as  $NN$ , it averages of determined weights are evaluated such as the following formulation:

$$O_i^3 = \frac{W_i}{\sum W_i} \quad \text{Eq. (13)}$$

Then in the next layer, the average weight values are multiplied to the related function such as below:

$$O_i^4 = \overline{W}_i f_i = \overline{W}_i (m_i X_1 + n_i X_2 + r_i) \quad \text{Eq. (14)}$$

Where, m, n, and r represent the resulting indexes.

At last, the fifth layer consists of the summation of previous layer outputs:

$$O_i^5 = Y = \sum_i \overline{W}_i f_i = \overline{W}_1 f_1 + \overline{W}_2 f_2 = \frac{\sum W_i f_i}{\sum W_i} \quad \text{Eq. (15)}$$

### 2.5 Particle swarm optimization (PSO)

The combination of random probability distribution approach and generation of the population constructed the particle swarm optimization algorithm. Eberhart et al. introduced the PSO algorithm that comes from the social behavior of birds and developed it to solve nonlinear function optimization problems (Kennedy 2010). This strategy has special similarities with other optimization approach such as genetic algorithm which is constructed base on random solution population. Each particle can be known as a probable solution of problem. A random population of particle created in search space to relate in optimum system.  $P_{best}$  is known as the best solution which can obtained from this strategy for a particle. Also  $g_{best}$  represents the global best solution determined by swarm. The particle move in the space by time iterations and the next iteration velocity is determined by using  $g_{best}$ ,  $P_{best}$  and current velocity (Eberhart and Kennedy 1995). The P'th particle can be determined as follow:

$$X_{pd}^{iter+1} = X_{pd}^{iter} + V_{pd}^{iter+1} \quad \text{Eq. (16)}$$

The particle velocity is updated by the following expression:

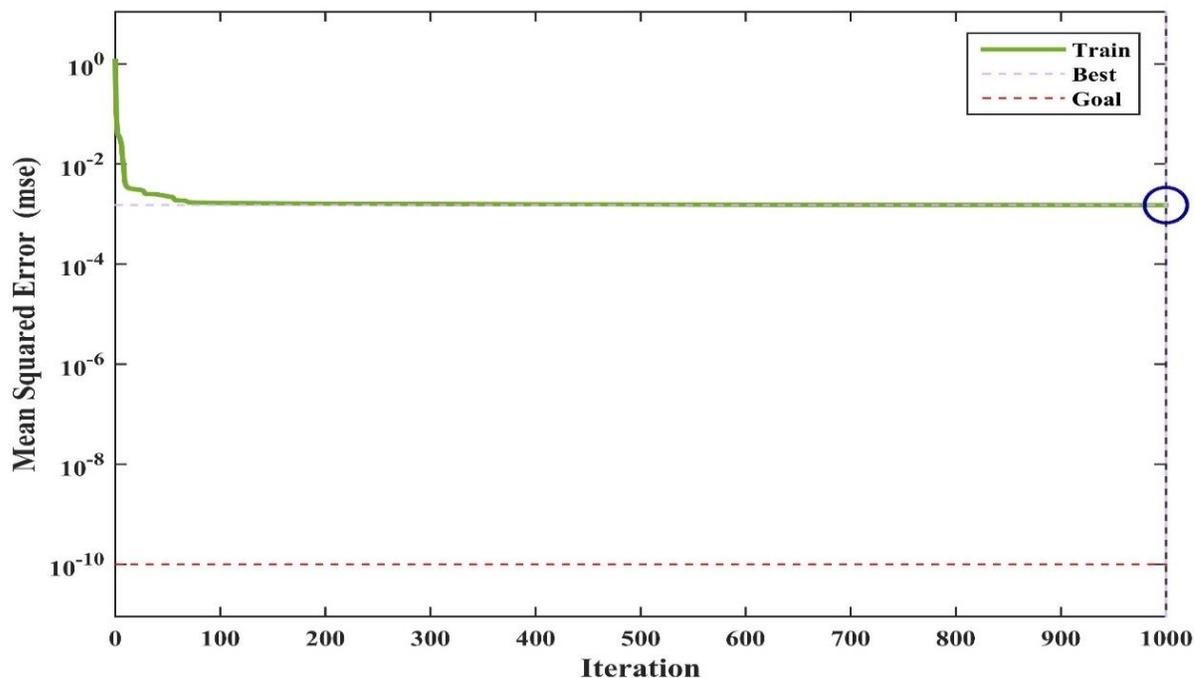
$$v_{id}(t+1) = wv_{id}(t) + c_1 r_1 (p_{best,id}(t) - X_{iid}(t)) + c_2 r_2 (g_{best,d}(t) - X_{id}(t)) \quad \text{Eq. (17)}$$

w, c, and r are inertia weight, learning rate and random number respectively (Haratipour, Baghban et al. 2017).

## 3 Results and discussion

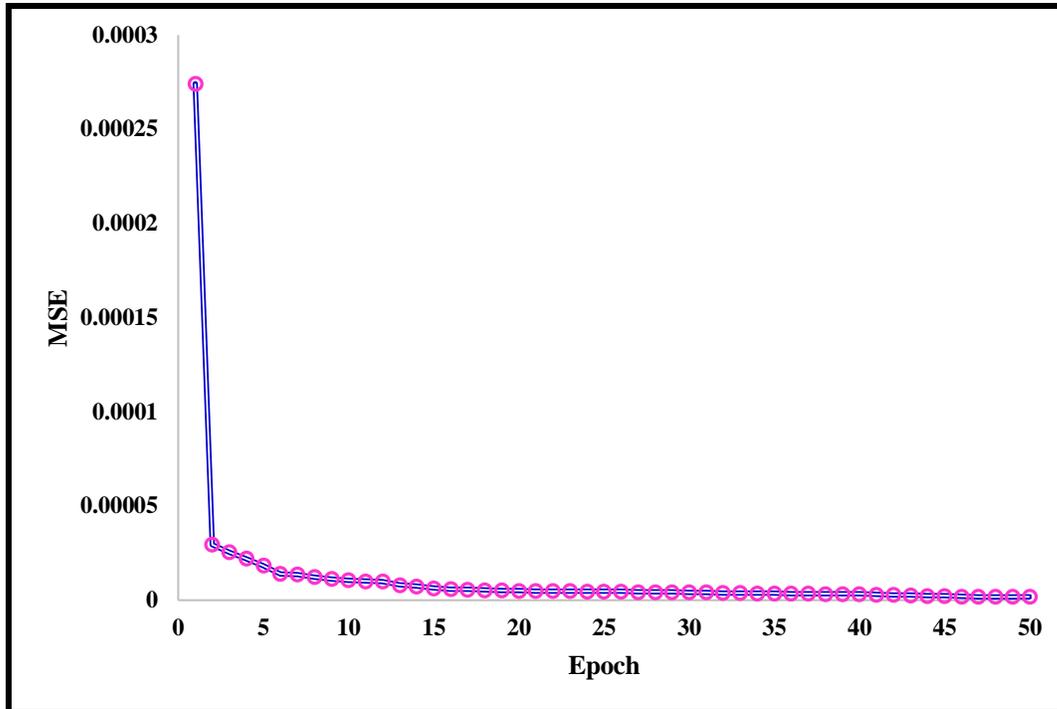
In the present study, the determined structure of MLP-ANN algorithm utilizes log-sigmoid and linear activation functions the hidden and output layers respectively. By utilization of trial and error, the optimum number of neurons in hidden layers is determined as 7 to reach the best

structure of MLP-ANN algorithm. The performance of Levenberg Marquardt training of MLP-ANN algorithm based on the mean square error is shown in **Fig. 2**.



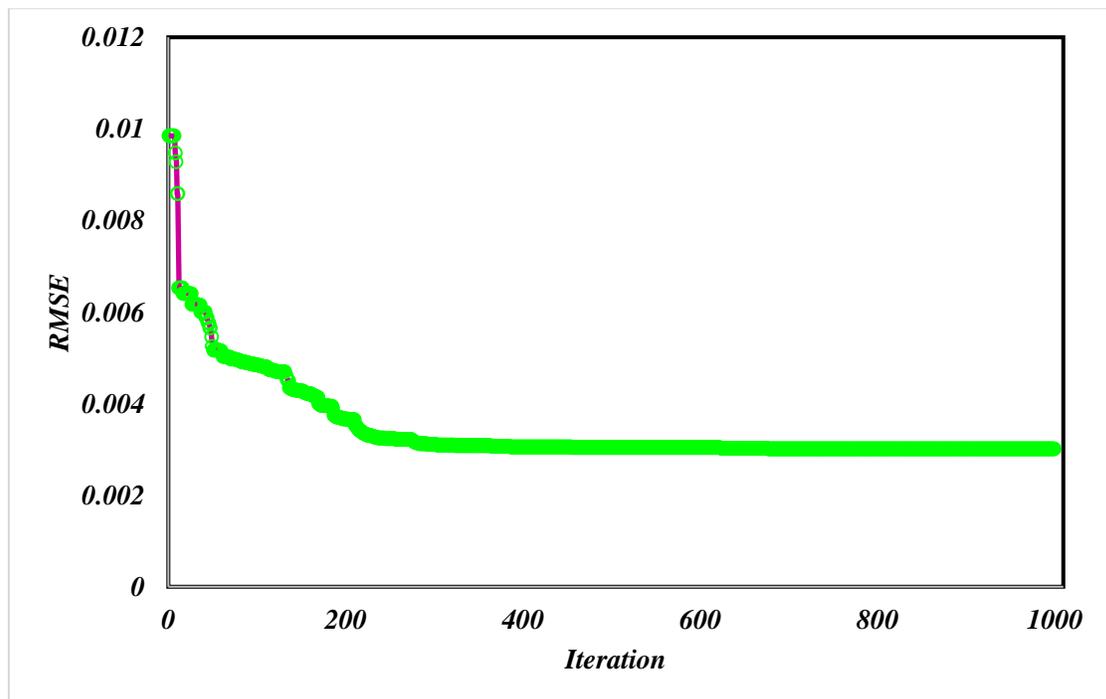
**Figure 2:** Trained MLP-ANN model by Levenberg Marquardt algorithm

In the RBF-ANN algorithm, the radial basis function (RBF) is utilized for hidden layers. According to information in the literature, the hidden layer neurons for RBF-ANN can be supposed one-tenth of training data points. The training process of RBF-ANN algorithm base on MSE has been reported in Fig. 3.



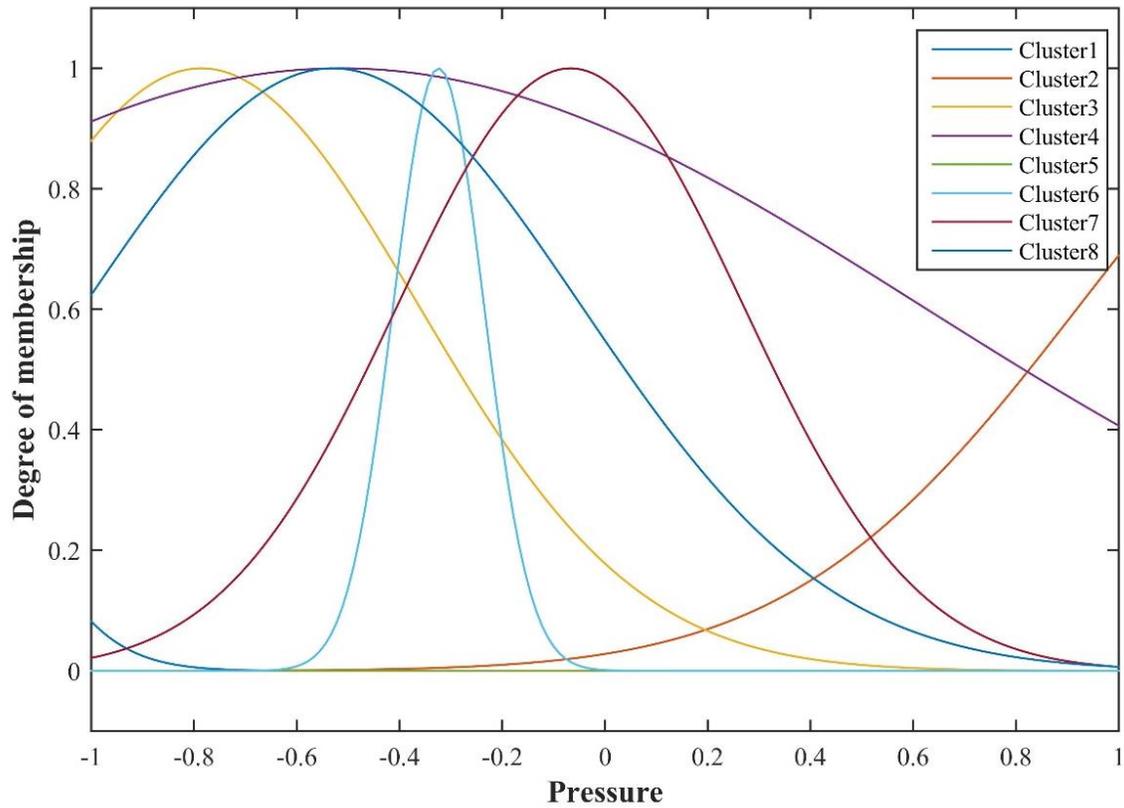
**Figure 3:** Trained RBF-ANN approach by Levenberg Marquardt algorithm

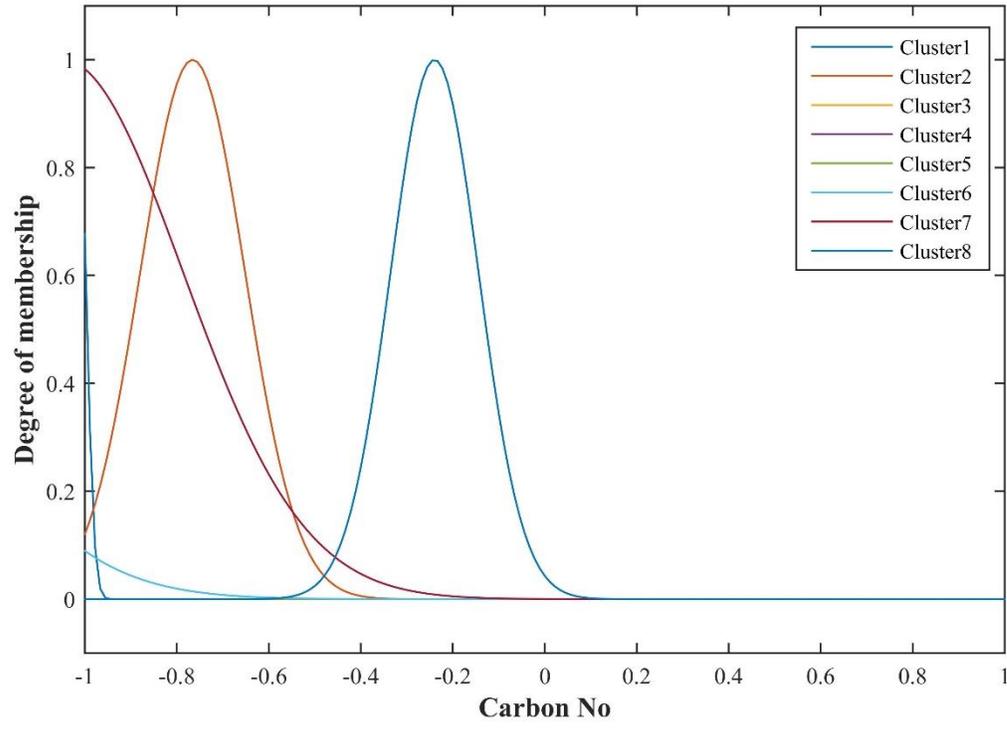
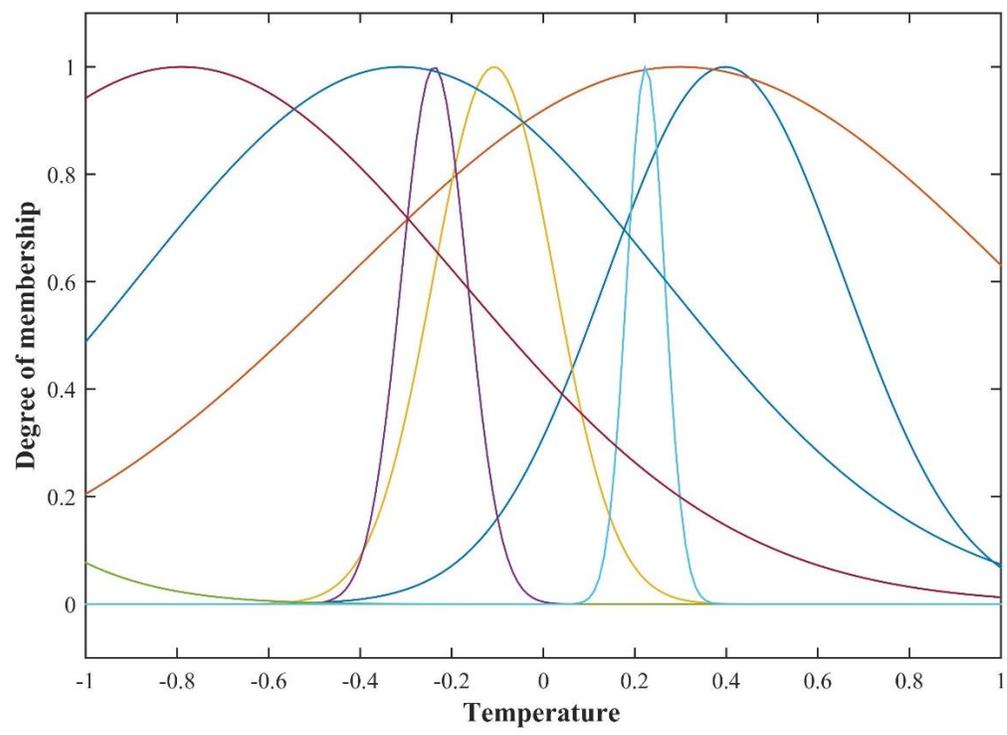
In this work, particle swarm optimization approach is applied to train the best structure of ANFIS algorithm. **Fig. 4** demonstrates the gained root mean squared error (RMSE) of estimated and experimental acid solubility values in training step.

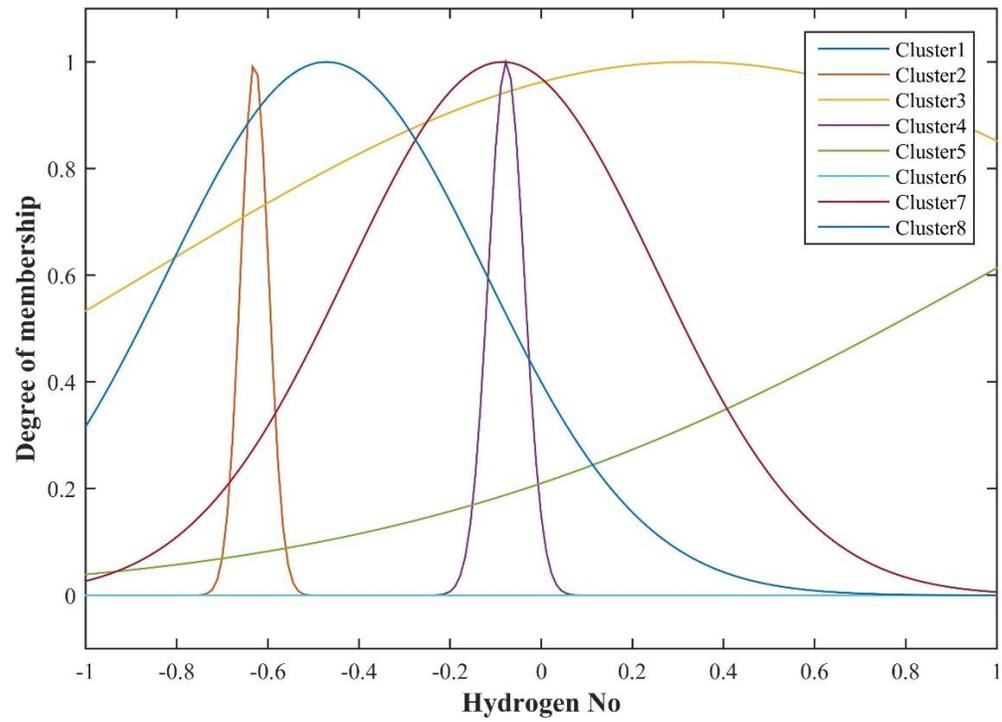
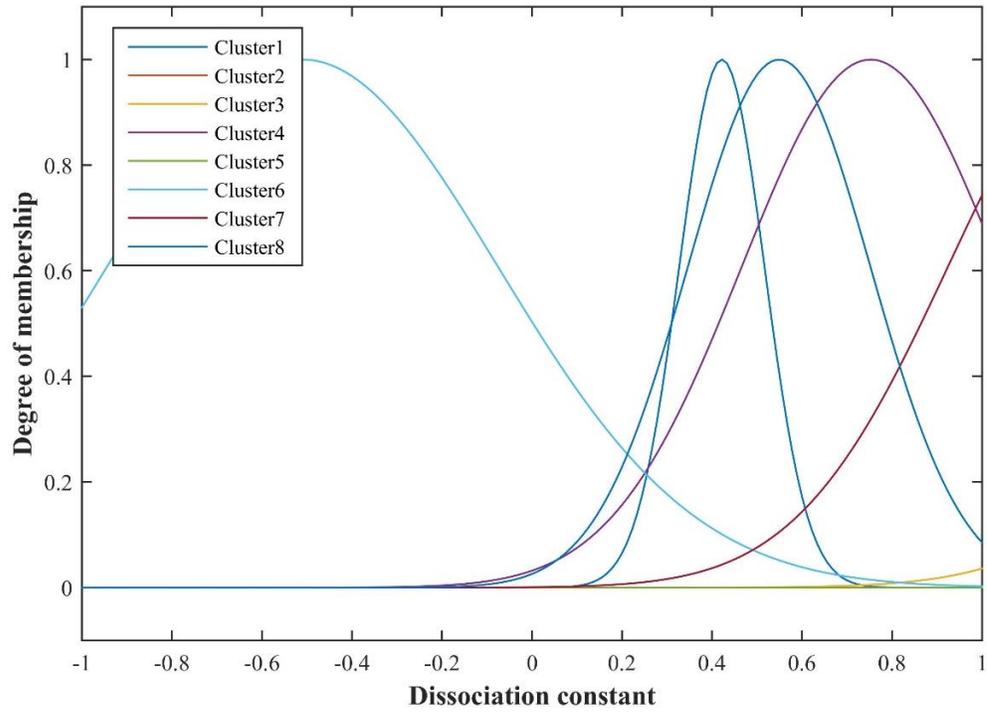


**Figure 4:** Performance of trained ANFIS model

The optimum structure of ANFIS can be recognized by the RMSE value of 0.003 after 1000 of iteration steps. Trained membership functions of proposed ANFIS model are also shown in **Fig. 5** for each cluster.







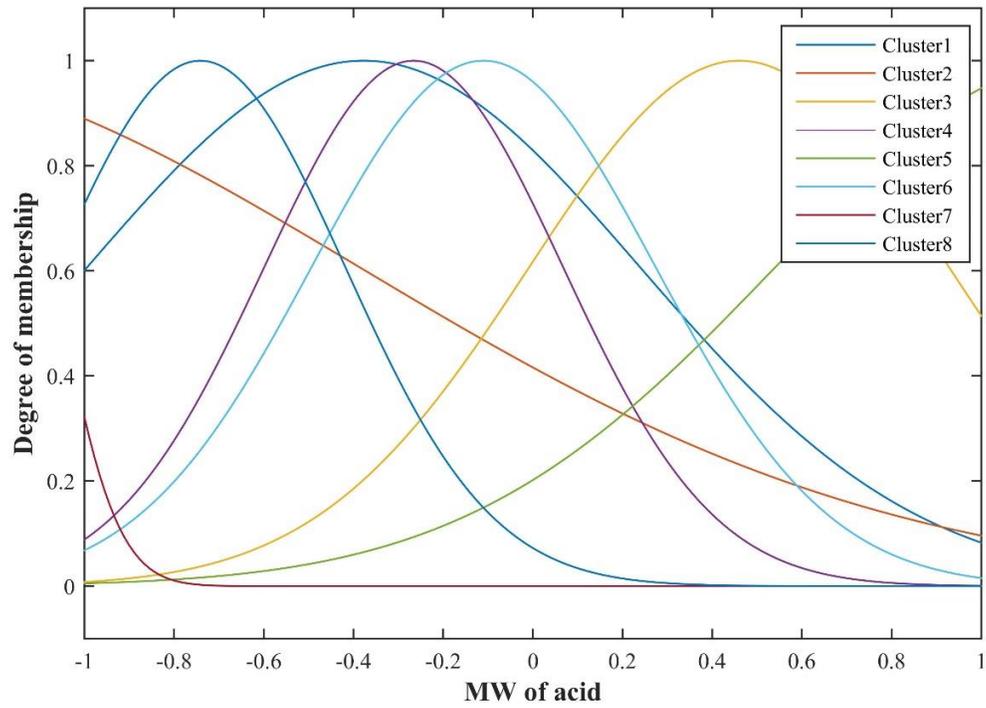


Figure 5: Trained membership function parameters

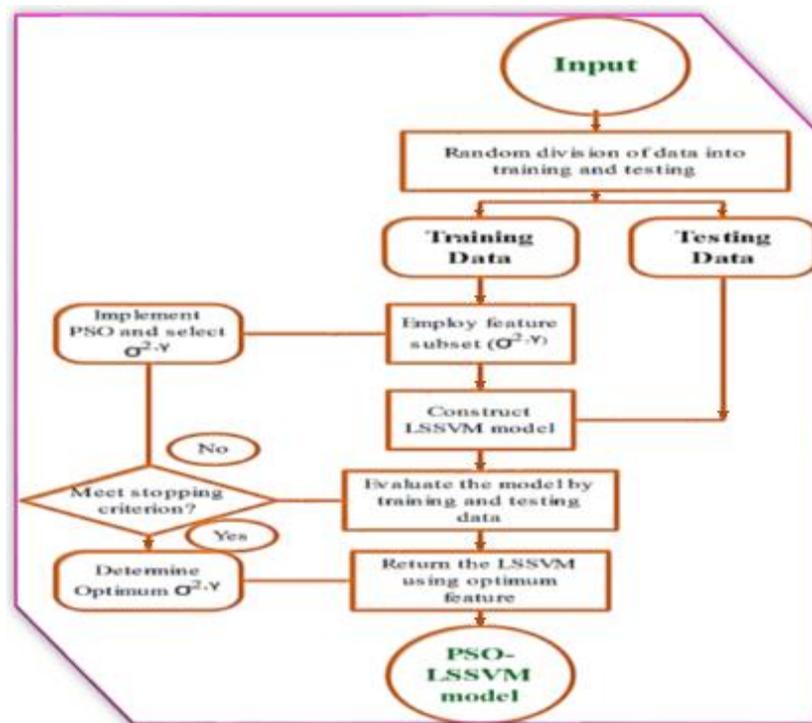


Figure 6: Schematic demonstration of trained LSSVM algorithm

The RBF kernel function due to its high degree of performance is utilized to construct the LSSVM algorithm. The LSSVM algorithm has two tuning parameters,  $\sigma^2$  and  $\gamma$  which are determined by utilizing PSO algorithm. The schematic demonstration of LSSVM algorithm is depicted in Fig. 6. The details of predicting models are summarized in Tab. 1. These details can be helpful in development of models for prediction of acid solubility in carbon dioxide.

**Table 1:** Details of proposed models

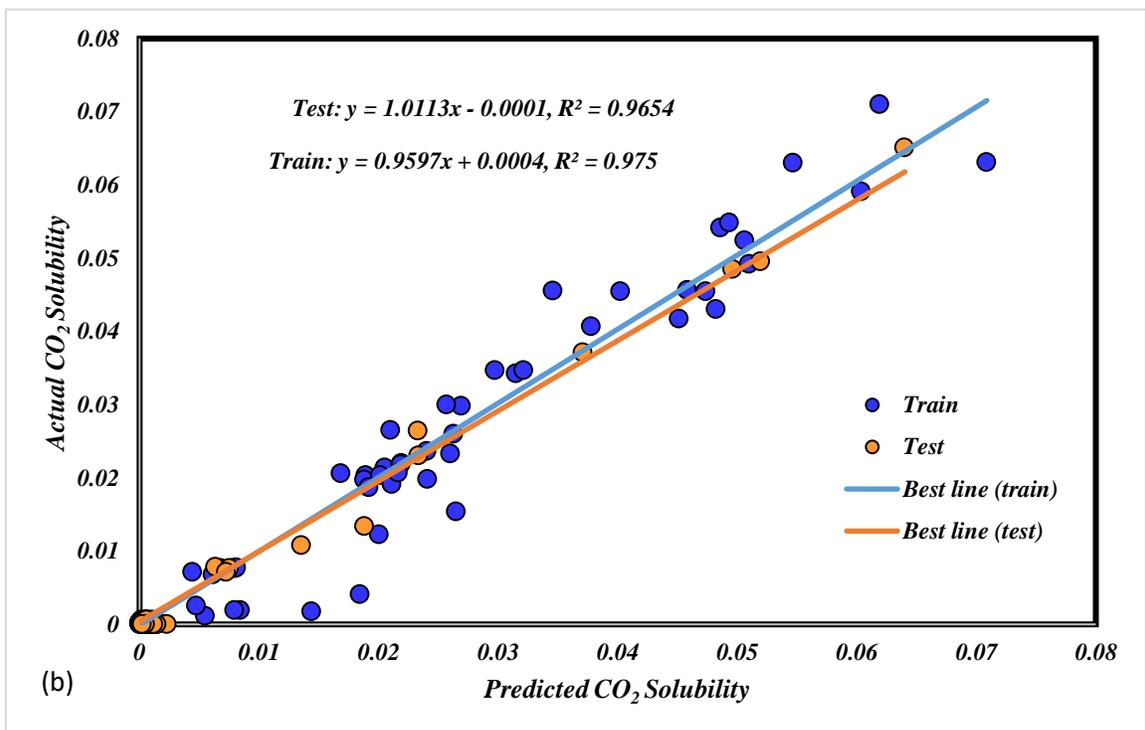
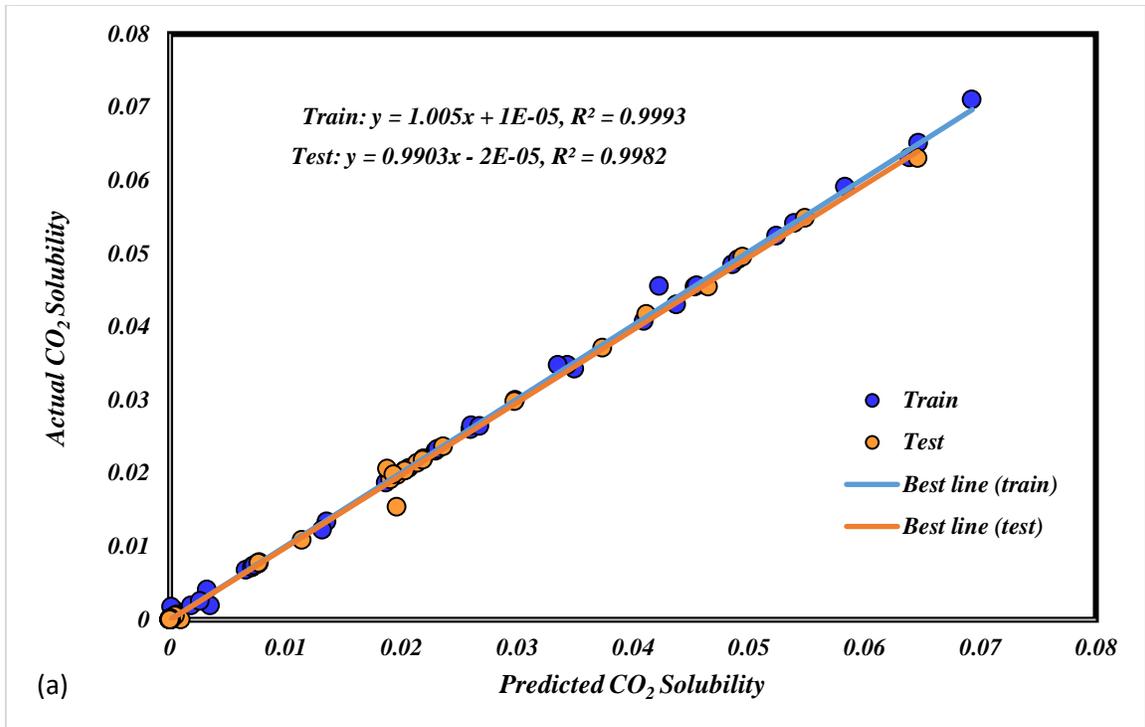
Type	comment/val ue	Type	comment/val ue
<b>LSSVM</b>		<b>ANFIS</b>	
<b>Kernel function</b>	RBF	Membership function	Gaussian
$\sigma^2$	0.80321	No. of membership function parameters	112
$\gamma$	12893.2264	No. of clusters	8
<b>Number of data utilized for training</b>	141	Number of data utilized for training	141
<b>Number of data utilized for testing</b>	47	Number of data utilized for testing	47
<b>Population size</b>	85	Population size	50
<b>Iteration</b>	1000	Iteration	1000
<b>C1</b>	1	C1	1
<b>C2</b>	2	C2	2
<b>MLP-ANN</b>		<b>RBF-ANN</b>	
<b>No. input neuron layer</b>	6	No. input neuron layer	6
<b>No. hidden neuron layer</b>	8	No. hidden neuron layer	50
<b>No. output neuron layer</b>	1	No. output neuron layer	1
<b>Hidden layer activation function</b>	Sigmoid	Hidden layer activation function	RBF
<b>output layer activation function</b>	linear	output layer activation function	linear
<b>Number of data utilized for training</b>	141	Number of data utilized for training	141

---

<b>Number of data utilized for testing</b>	47	Number of data utilized for testing	47
<b>Number of max iteration</b>	1500	Number of max iteration	50

---

In order to show the performance of proposed models in prediction of solubility of different acids, regression plots of RBF-ANN, MLP-ANN, ANFIS and LSSVM algorithms are depicted in Fig. 7 to compare the determined and actual solubility values. Based on these plots, the surprising fits for the predicting algorithms are obtained. Also, the predicted acid solubility data for proposed models are demonstrated along with the corresponding actual acid solubility values in Fig. S1. It can be observed that the model's output solubility values have excellent agreement with actual solubility values. Another graphical evaluation method is demonstration of relative error between predicted and experimental acid solubility in supercritical carbon dioxide. Fig. S2 shows the percentage of absolute error for the different predicting algorithm. The percentages of absolute error place under 1.5 percent for all developed algorithms, which expresses the acceptable degree of accuracy in prediction of acid solubility.



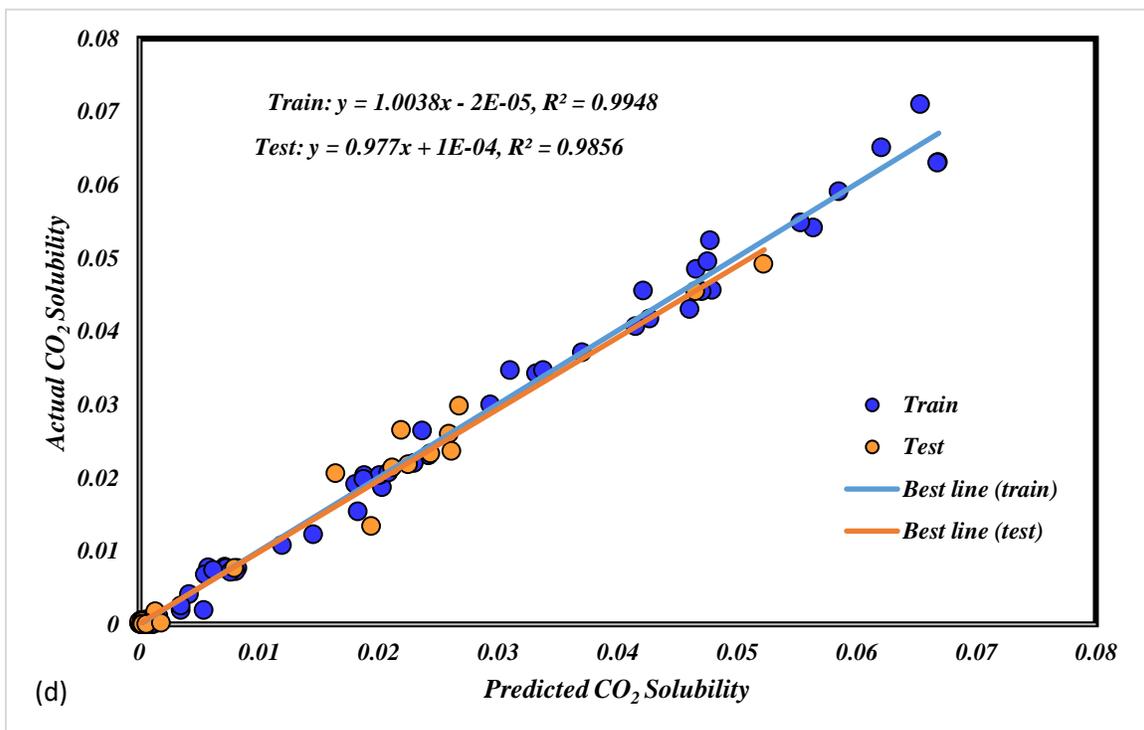
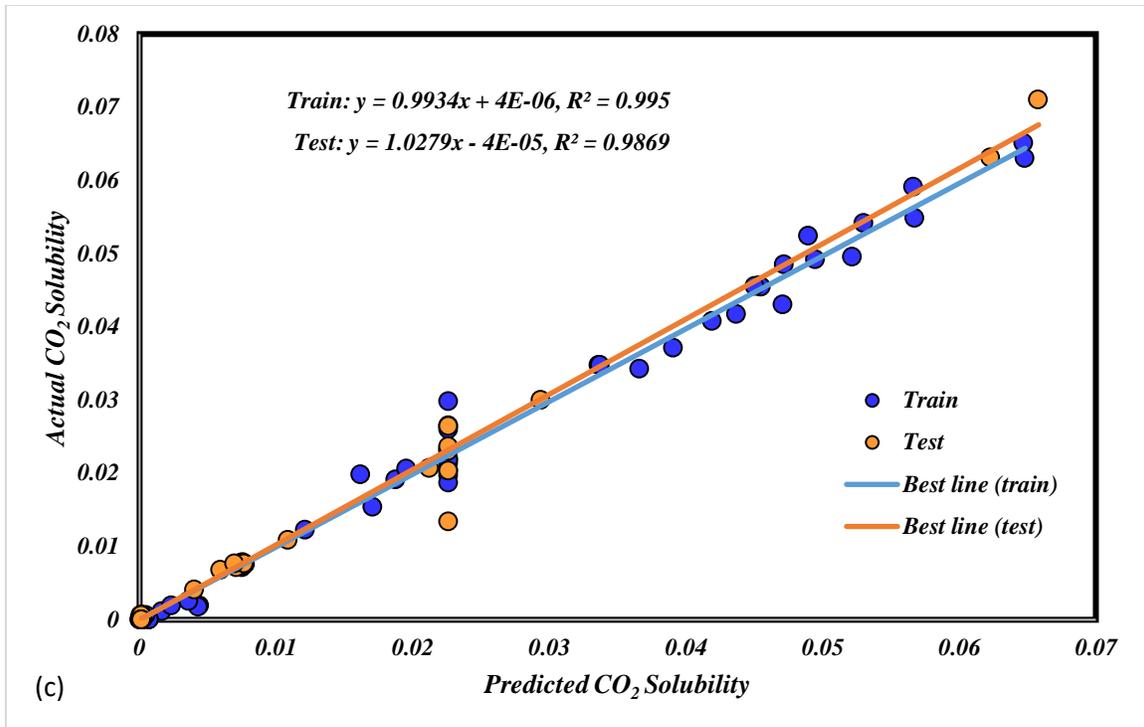


Figure 7: Regression plots obtained for different models

Furthermore, in order to clarify the performance of predicting algorithms, the statistical analysis is required so the coefficients of determination ( $R^2$ ), average absolute deviation (AAD), Mean squared errors (MSEs) and Standard deviations (STDs) are determined such as following:

$$R^2 = 1 - \frac{\sum_{i=1}^N (X_i^{\text{actual}} - X_i^{\text{predicted}})^2}{\sum_{i=1}^N (X_i^{\text{actual}} - \bar{X}^{\text{actual}})^2} \quad \text{Eq. (18)}$$

$$AAD = \frac{1}{N} \sum_{i=1}^N |X_i^{\text{predicted}} - X_i^{\text{actual}}| \quad \text{Eq. (19)}$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i^{\text{actual}} - X_i^{\text{predicted}})^2 \quad \text{Eq. (20)}$$

$$STD_{\text{error}} = \left( \frac{1}{N-1} \sum_{i=1}^N (\text{error} - \overline{\text{error}}) \right)^{0.5} \quad \text{Eq. (21)}$$

The  $R^2$ , AD, MSE and STD values of different algorithms are summarized in **Tab. 2**. According to these results, the LSSVM model has the greatest ability in forecasting acid solubility. The determined  $R^2$  values for LSSVM is equal to 0.998 and 0.999 in train and test set, respectively. Furthermore it's RMSE, MSE and AAD parameters are 0.000527, 2.77875E-07, and 0.0179, respectively. According to these analyses LSSVM algorithm is known as the best predictor for prediction of solubility of different acids.

**Table 2:** Statistical analyses of models

Model	Set	MSE	RMSE	$R^2$	STD	AAD (%)
<b>LSSVM</b>	Train	5.72159E-07	0.000756	0.998	0.0007	0.0269
	Test	1.7978E-07	0.000424	0.999	0.0004	0.0149
	Total	2.77875E-07	0.000527	0.999	0.0005	0.0179
<b>ANFIS</b>	Train	5.79633E-06	0.002408	0.975	0.0022	0.1093
	Test	1.00976E-05	0.003178	0.965	0.0027	0.1677
	Total	9.02227E-06	0.003004	0.967	0.0026	0.1531
<b>MLP-ANN</b>	Train	3.23782E-06	0.001799	0.987	0.0017	0.0756
	Test	1.44839E-06	0.001203	0.995	0.0010	0.0600
	Total	1.89575E-06	0.001377	0.993	0.0012	0.0639
<b>RBF-ANN</b>	Train	2.33037E-06	0.001527	0.986	0.0013	0.0827
	Test	1.61993E-06	0.001273	0.995	0.0010	0.0779
	Total	1.79754E-06	0.001341	0.993	0.0011	0.0791

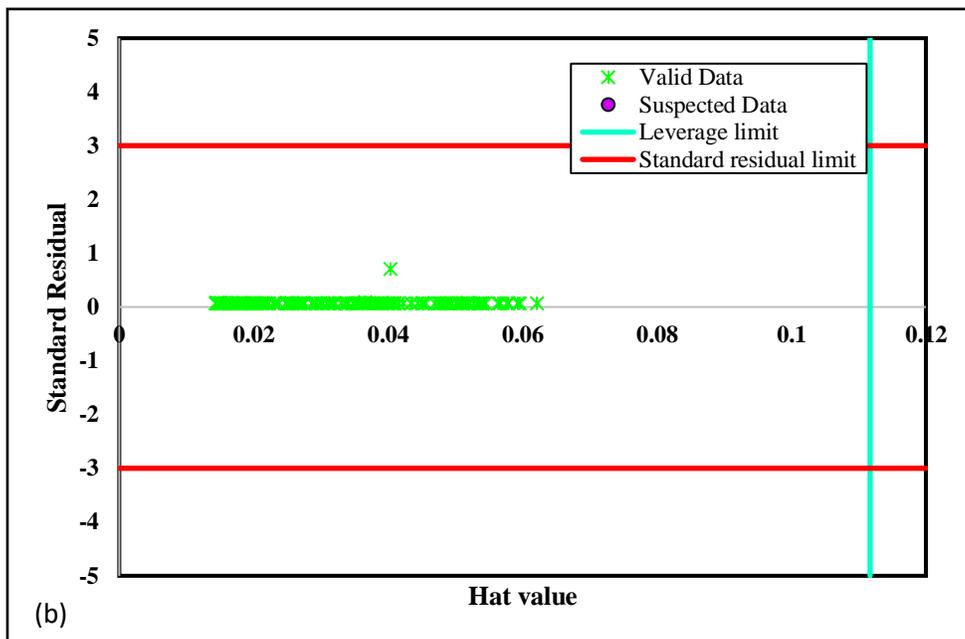
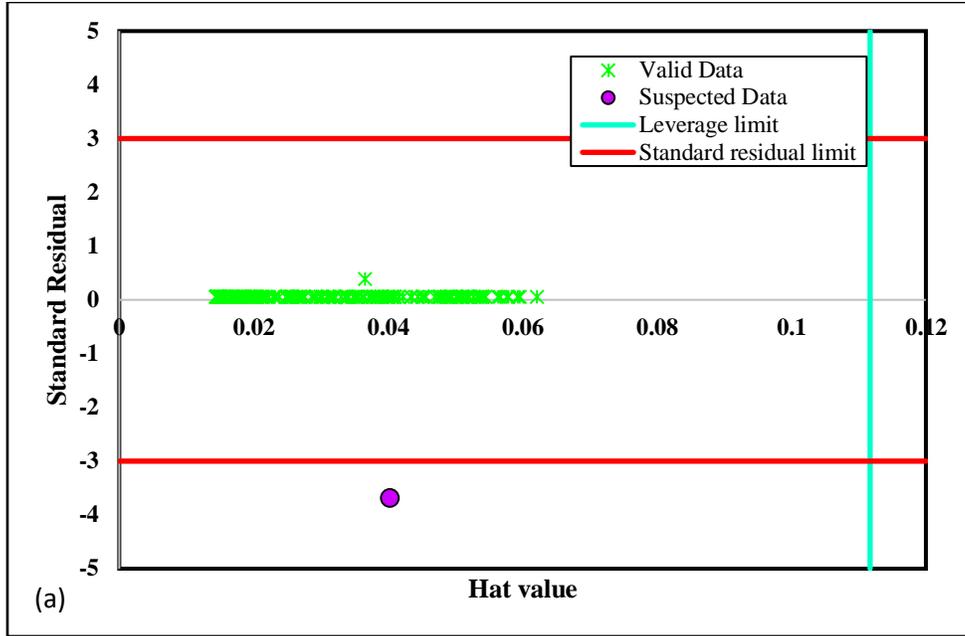
In addition to previous statistical indexes, there is another statistical approach to evaluate the reliability and accuracy of predicting algorithm, which called Leverage method. The mentioned approach consists of some statistical concepts such as model residuals, Hat matrix, and Williams plot which are used for detection of suspected and outlier data. There is more description of Leverage method in the literature (Rousseeuw and Leroy 2005). In this method, the residuals are estimated and inputs are utilized to build a matrix called Hat matrix such as follow:

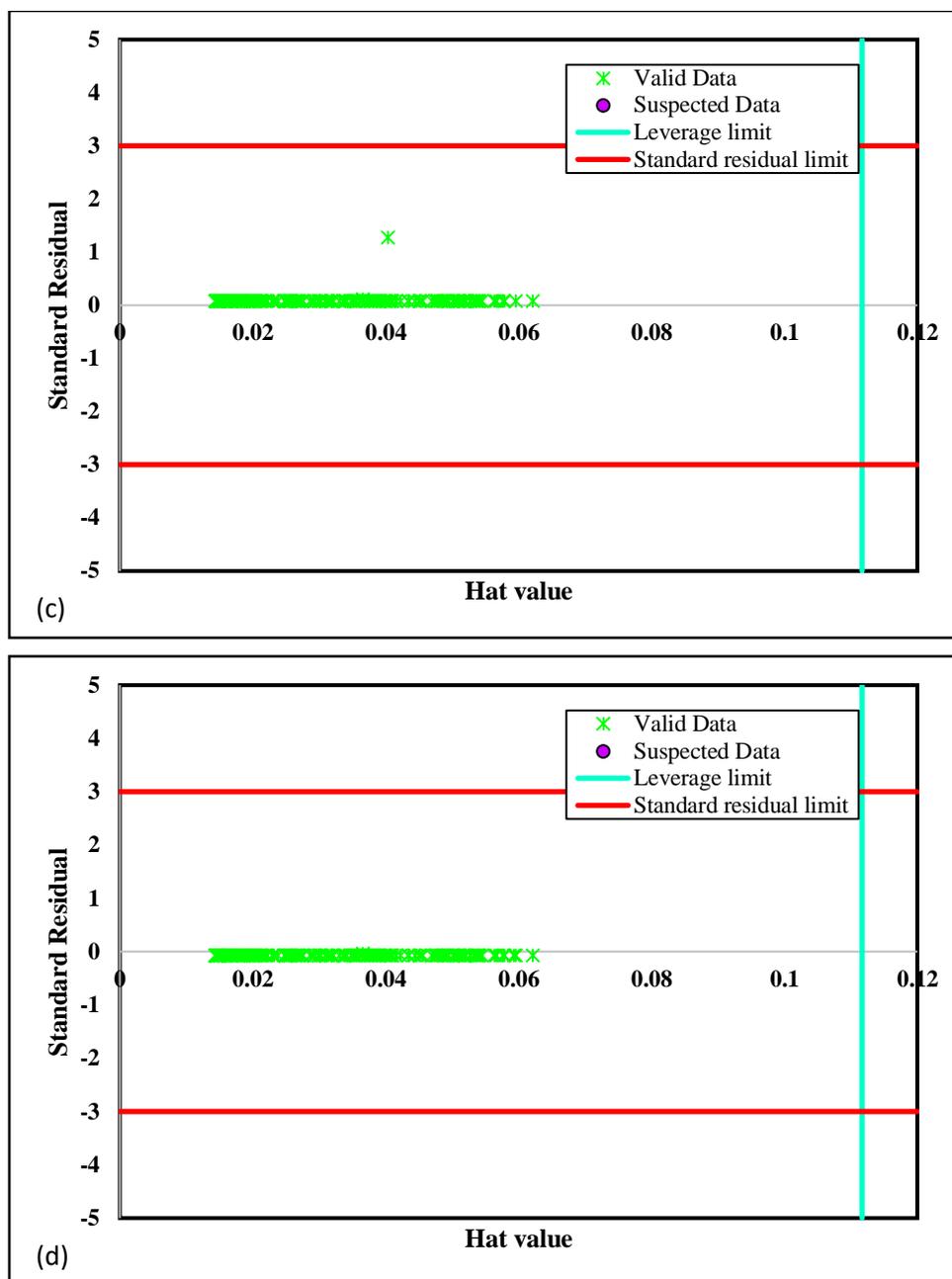
$$H = X(X^T X)^{-1} X^T \quad \text{Eq. (22)}$$

Where  $X$  is the  $m \times n$  matrix which  $n$  and  $m$  are the numbers of model parameters and samples respectively.

**Fig. 8** illustrates the William plot for the proposed models. As shown in this figure, the most of data points are in the range of leverage limit of residuals for -3 to 3. The leverage limit is formulated such as following:

$$H^* = 3(n + 1)/m \quad \text{Eq. (23)}$$



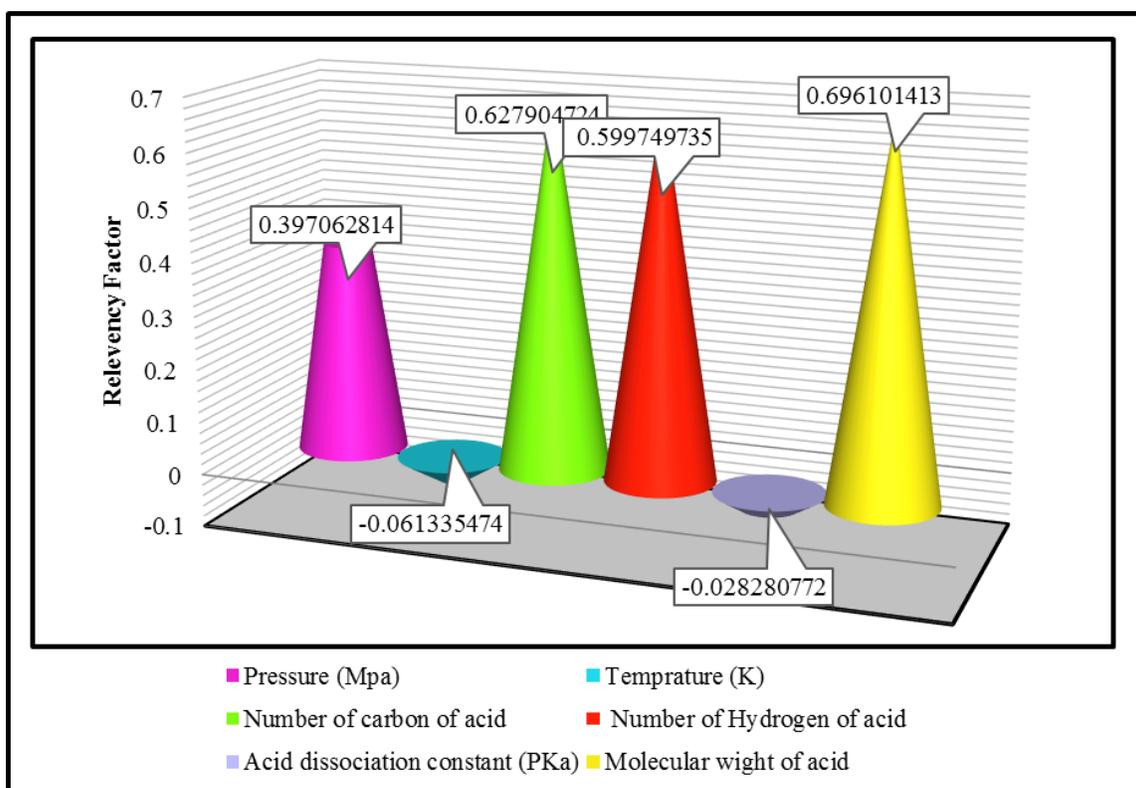


**Figure 8:** Absolute deviation plots for (a) LSSVM, (b) ANFIS, (c) MLP-ANN, and (d) RBF-ANN

Another method to investigate the validity of the models is a parametric analysis of solubility. To this end, the Relevancy index is introduced to investigate the impact of inputs on acid solubility. The Relevancy index is determined such as following (Zarei, Razavi et al. 2019):

$$r = \frac{\sum_{i=1}^n (X_{k,i} - \bar{X}_k)(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_{k,i} - \bar{X}_k)^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad \text{Eq. (24)}$$

where  $Y_i$ ,  $\bar{Y}$ ,  $X_{k,i}$  and  $\bar{X}_k$  are the 'i' th output, output average, kth of input and average of input. The Relevancy index absolute value represent the effectiveness of the parameters on acid solubility. As shown in **Fig. 9**, the molecular weight of acid has the most Relevancy factor between different input parameters so this parameter is known as the most effective parameters on acid solubility in supercritical carbon dioxide. Moreover, acid dissociation constant has the least effect on acid solubility. This figure illustrates that as number of carbon and hydrogen of acid, molecular weight and pressure increase, acid solubility in carbon dioxide increases. On the other hand, increasing acid dissociation constant and temperature caused drop in solubility of acid in carbon dioxide.



**Figure 9:** Sensitivity analysis of investigated variables

#### 4 Conclusions

In this paper, we have applied RBF-ANN, MLP-ANN, ANFIS-PSO and LSSVM algorithms to determine the different acids solubility values in supercritical carbon dioxide in terms of pressure,

temperature, and different acid structure based on a reliable databank which gathered from the literature. These predicting approaches can forecast acid solubility in the wide range of operating conditions. To prove the aforementioned acclaim, different statistical and graphical evaluations have been performed in the previous section. According to the obtained results from comparisons, the LSSVM model has the best performance respect to the others and ANFIS algorithm has the least of accuracy in this prediction. Also, the results of sensitivity analysis identify the molecular weight of the acid parameter is the most effective factor in solubility of acids in supercritical carbon dioxide. Based on these comprehensive investigations this manuscript has great potential and ability to help the researchers in their future works.

## Nomenclature

<b>ANFIS</b>	Adaptive neuro-fuzzy inference system
<b>LSSVM</b>	Least squares support vector machine
<b>RBF-ANN</b>	Radial basis function artificial neural network
<b>MLP-ANN</b>	Multi-layer Perceptron artificial neural network
<b>PSO</b>	Particle swarm optimization
$\varphi(x)$	nonlinear function
$\omega$	weight
<b>b</b>	bias
$\gamma$	regularization parameter
$e_k$	support value
<b>K</b>	kernel function
<b>Z</b>	Gaussian parameter
$\sigma$	Gaussian parameter
<b>m</b>	One of the resulting index of ANFIS
<b>n</b>	One of the resulting index of ANFIS
<b>r</b>	One of the resulting index of ANFIS
<b>W</b>	inertia weight
<b>c</b>	learning rate
<b>R<sup>2</sup></b>	coefficient of determination
<b>AAD</b>	average absolute deviation
<b>MSE</b>	Mean squared error
<b>STD</b>	Standard deviation
<b>H</b>	Hat matrix
<b>H<sup>*</sup></b>	The leverage limit

## Supplementary content

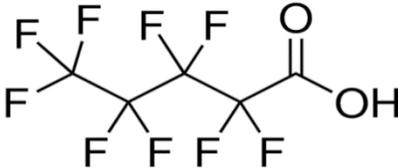
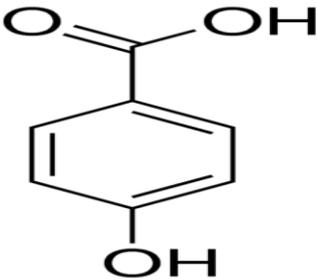
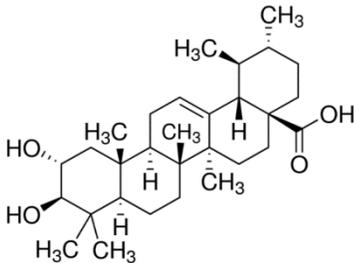
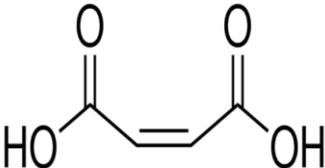
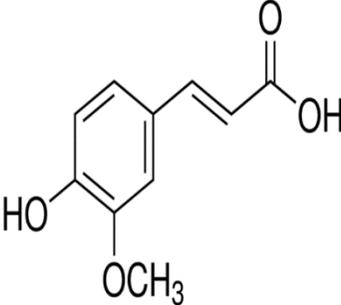
**Table S1.** Experimental data which are used in this study

Acid name	Pressure	Temperature (K)	Acid dissociation constant (PKa)	solubility (mol/mol)	No of data points	References
<b>Perfluoropentanoic acid</b>	10-26.2	314-334	0.52	0.0134-0.0298	17	(Dartiguelon gue, Leybros et al. 2016)
<b>o-Hydroxybenzoic Acid</b>	8.11-20.26	308.15-328.15	4.06	0.000007-0.000624	49	(Gurdial and Foster 1991)
<b>Corosolic Acid</b>	8.0-30	308.15-333.15	4.7	3.28*10 <sup>-11</sup> - 0.071	40	(Kumoro 2011)
<b>Maleic Acid</b>	7.0-30	318.15-348.15	1.83	0.000013-0.0005917	21	(Sahihi, Ghaziaskar et al. 2010)
<b>Ferulic Acid</b>	12.0-28	301.15-333.15	4.38	0.00000155-0.0000118	18	(Sovova, Zarevucka et al. 2001)
<b>Azelaic Acid</b>	10.0-30	313.15-333.15	4.84	0.00000042-0.00001012	14	(Sparks, Hernandez et al. 2007)
<b>Nonanoic Acid</b>	10.0-30	313.15-333.15	4.96	0.00013-0.00782	14	(Sparks, Estévez et al. 2008)
<b>p-aminobanzoic acid</b>	8.0-21	308-328.0	4.78	0.000001302-0.000006452	15	(Tian, Jin et al. 2007)
					Total=	
					188	

**Table S2.** Average of experimental data which are used in this study

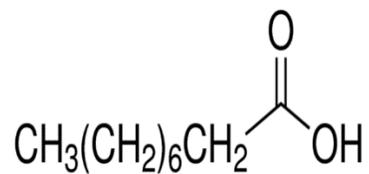
Acid name	Pressure (Mpa)	Temperature (K)	solubility(mol/mol)
Perfluoropentanoic acid	17.37058824	324	0.022118
o-Hydroxybenzoic Acid	13.84040816	316.6193878	0.000238
Corosolic Acid	18.2	319.4	0.029932
Maleic Acid	16.42857143	333.15	0.000173
Ferulic Acid	19.83333333	319.4833333	5.37E-06
Azelaic Acid	20	323.15	3.92E-06
Nonanoic (Pelargonic) Acid	20	323.15	0.006548
p-aminobanzoic acid	14	318	3.82E-06

**Table S3.** Details of acids which are utilized in this investigation.

Acid name	structure	Empirical Formula or linear formula	Molecular weight gr/mole
Perfluoropentanoic acid		$\text{CF}_3(\text{CF}_2)_3\text{COOH}$	264.05
o-Hydroxybenzoic Acid		$\text{HOC}_6\text{H}_4\text{CO}_2\text{H}$	138.12
Corosolic Acid		$\text{C}_{30}\text{H}_{48}\text{O}_4$	472.70
Maleic Acid		$\text{HO}_2\text{CCH}=\text{CHCO}_2\text{H}$	116.07
Ferulic Acid		$\text{HOC}_6\text{H}_3(\text{OCH}_3)\text{CH}=\text{CHCO}_2\text{H}$	194.18
Azelaic Acid		$\text{HO}_2\text{C}(\text{CH}_2)_7\text{CO}_2\text{H}$	188.22

---

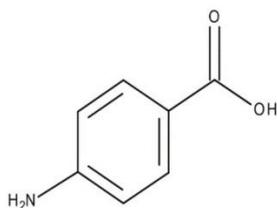
**Nonanoic (Sparks,  
Estévez et al. 2008)  
Acid**



$\text{CH}_3(\text{CH}_2)_7\text{COOH}$

158.24

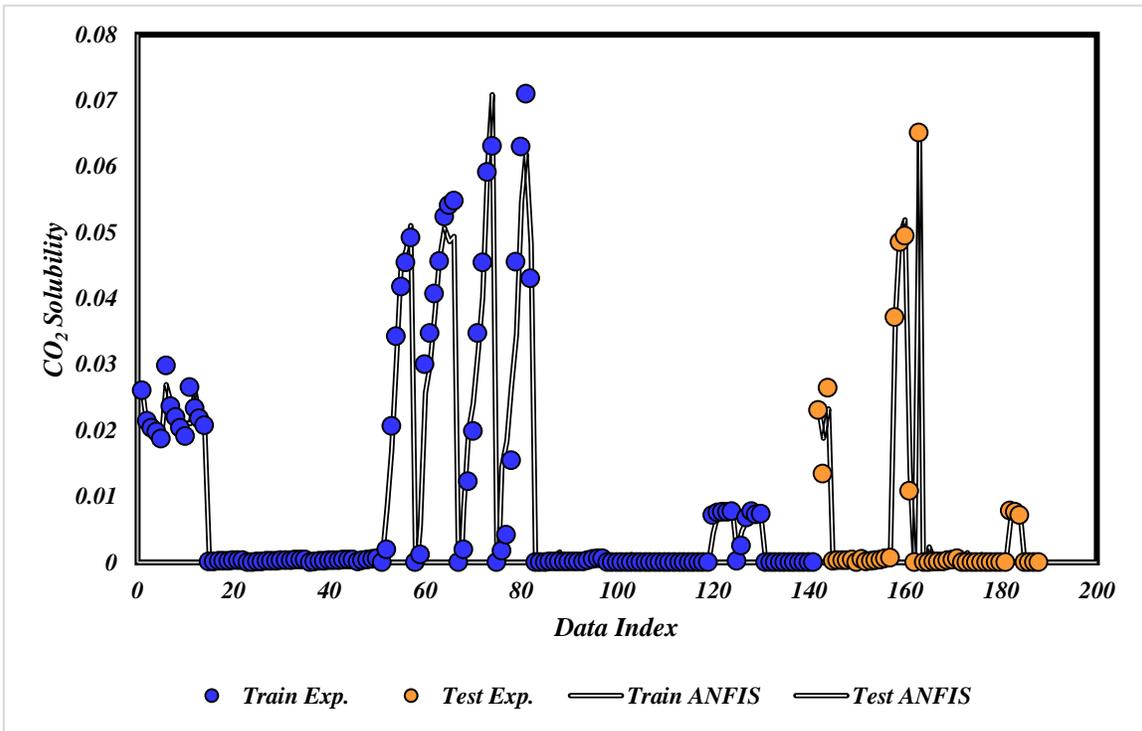
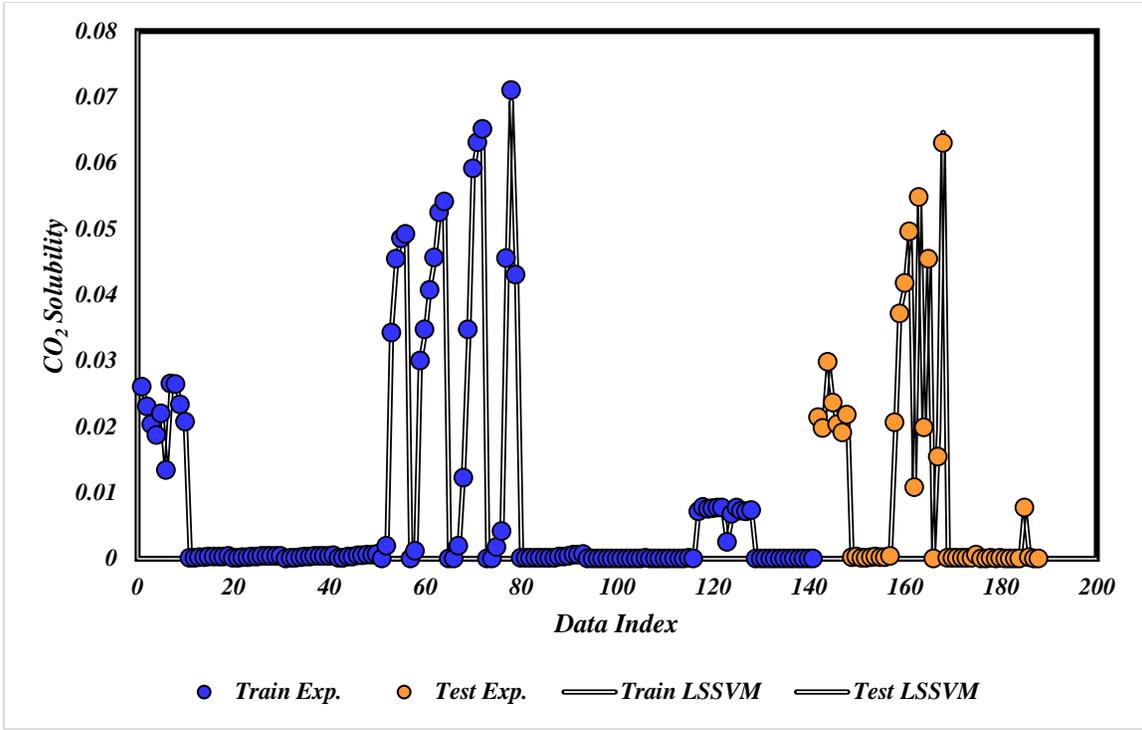
**p-aminobenzoic acid**

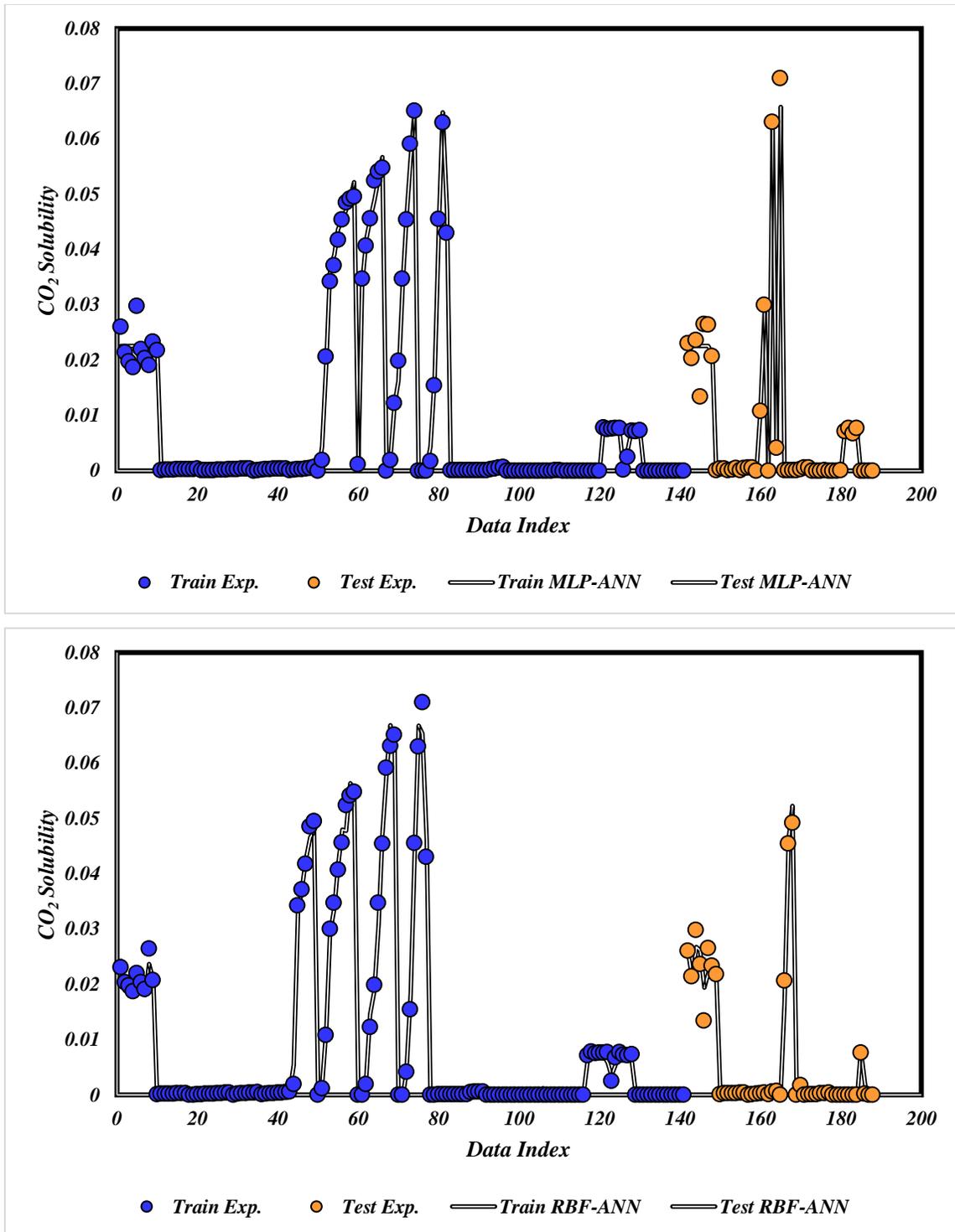


$\text{C}_7\text{H}_7\text{NO}_2$

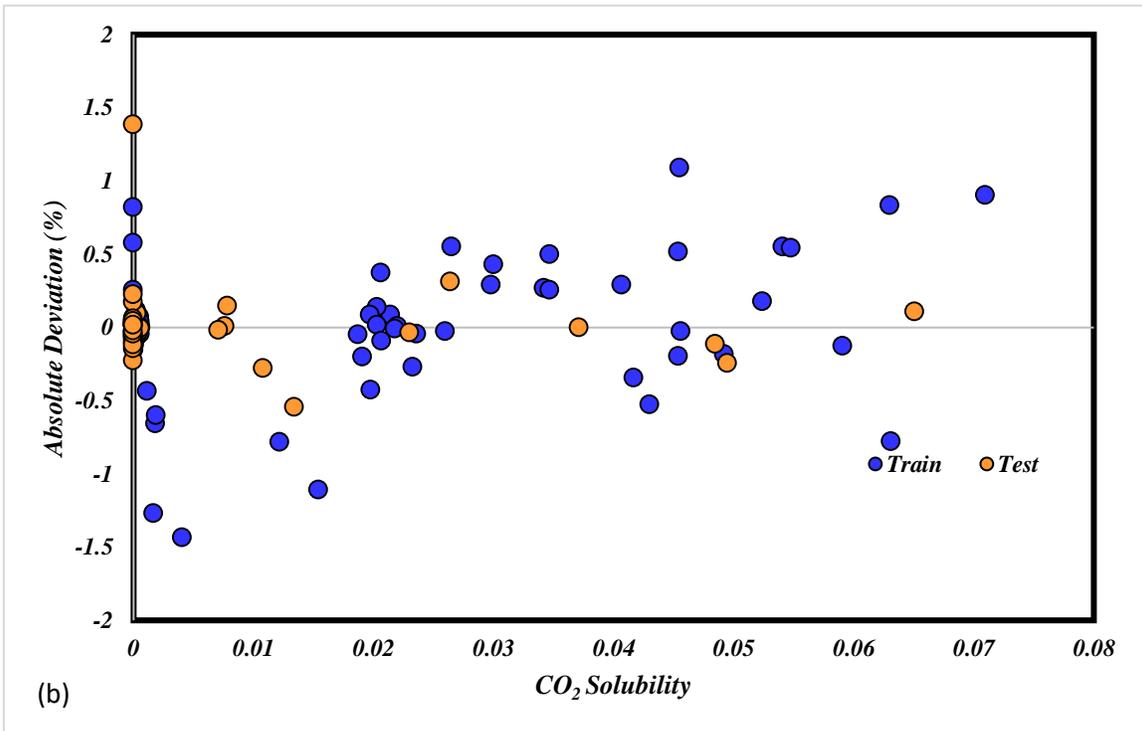
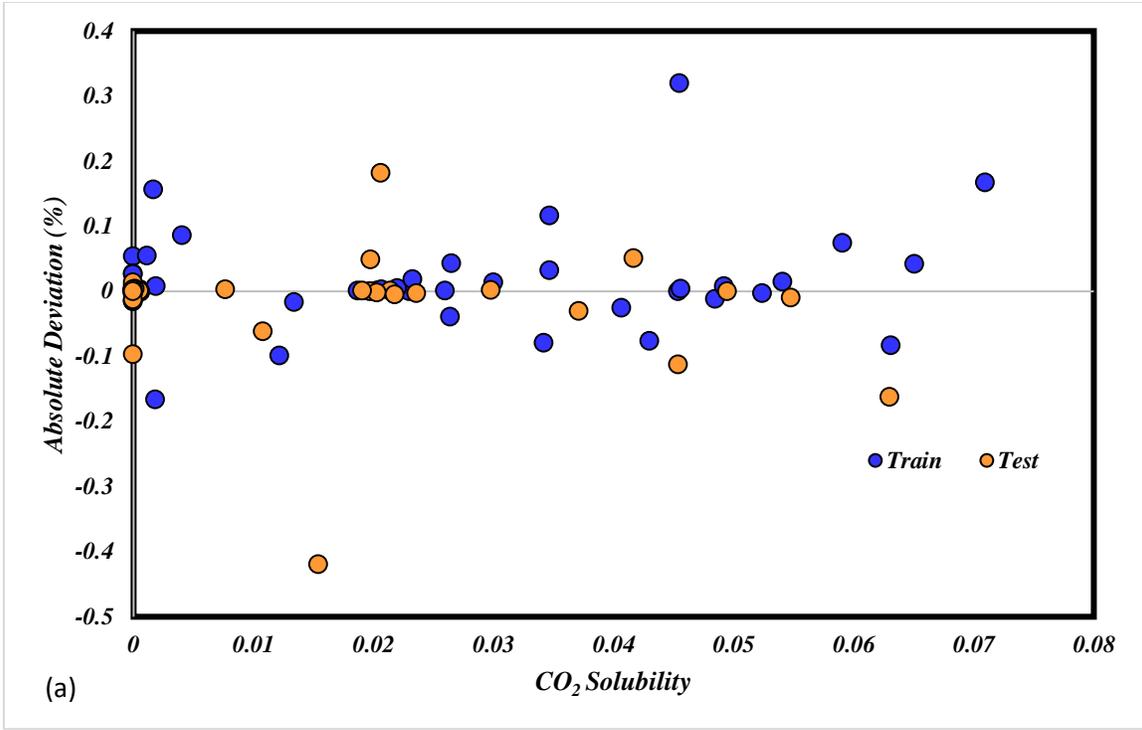
137.14

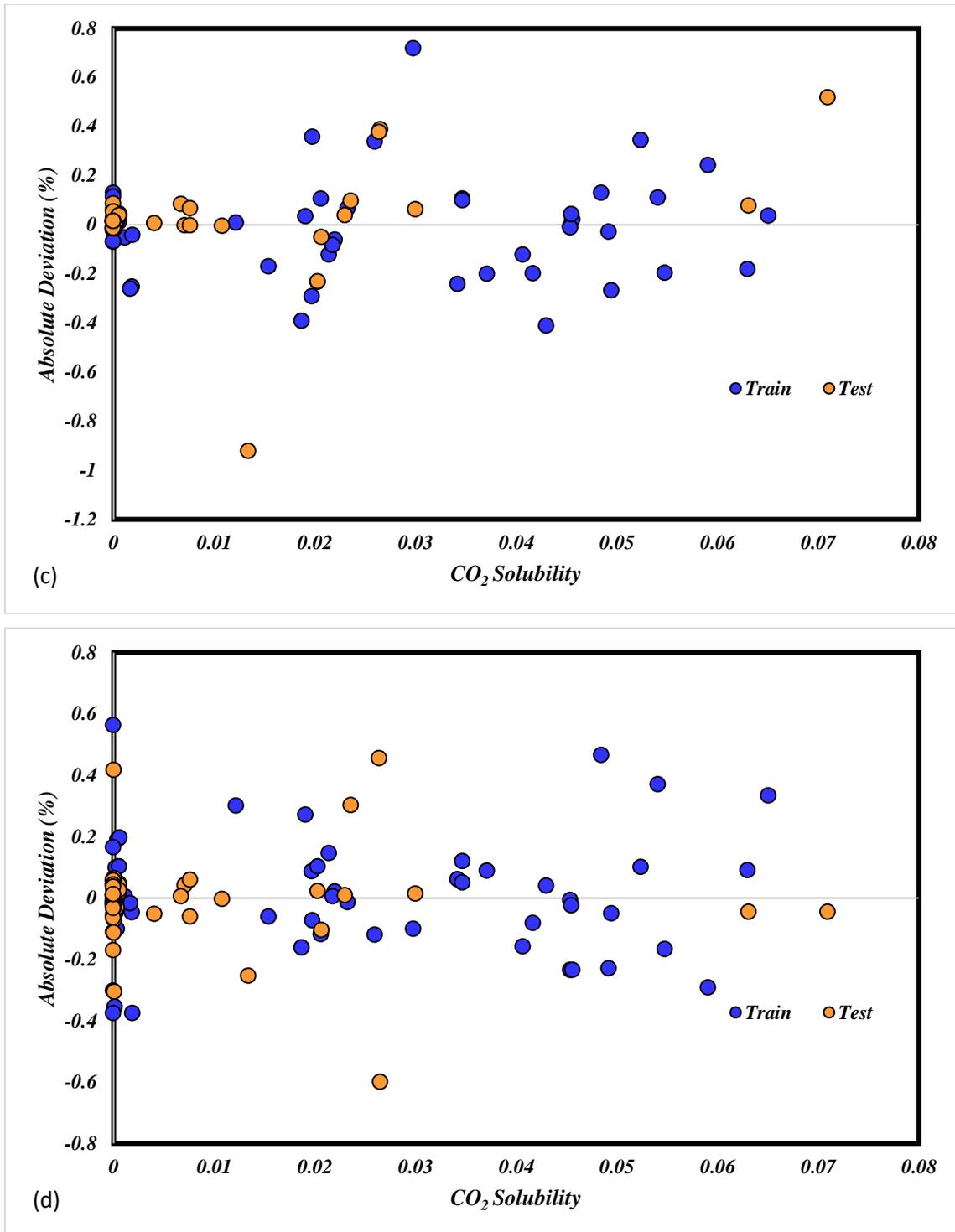
---





**Figure S1:** Experimental and predicted solubility of CO<sub>2</sub> by the proposed models





**Figure S2:** Absolut deviation plots for (a) LSSVM, (b) ANFIS, (c) MLP-ANN, and (d) RBF-ANN

## Acknowledgement

This publication has been supported by the Project: "Support of research and development activities of the J. Selye University in the field of Digital Slovakia and creative industry" of the Research & Innovation Operational Programme (ITMS code: NFP313010T504) co-funded by the European Regional Development Fund.

## References

- Abdi-Khanghah, M., A. Bemani, Z. Naserzadeh and Z. Zhang** (2018). "Prediction of solubility of N-alkanes in supercritical CO<sub>2</sub> using RBF-ANN and MLP-ANN." Journal of CO<sub>2</sub> Utilization **25**: 108-119.
- Afshar, M., A. Gholami and M. Asoodeh** (2014). "Genetic optimization of neural network and fuzzy logic for oil bubble point pressure modeling." Korean Journal of Chemical Engineering **31**(3): 496-502.
- Ahangari, K., S. R. Moeinossadat and D. Behnia** (2015). "Estimation of tunnelling-induced settlement by modern intelligent methods." Soils and Foundations **55**(4): 737-748.
- Ahmadi, M. H., A. Baghban, E. Salwana, M. Sadeghzadeh, M. Zamen, S. Shamshirband and R. Kumar** (2019). "Machine Learning Prediction Models of Electrical Efficiency of Photovoltaic-Thermal Collectors."
- Anitescu, C., E. Atroshchenko, N. Alajlan and T. Rabczuk** (2019). "Artificial neural network methods for the solution of second order boundary value problems." Computers, Materials & Continua **59**(1): 345-359.
- Baghban, A., M. Bahadori, A. S. Lemraski and A. Bahadori** (2016). "Prediction of solubility of ammonia in liquid electrolytes using least square support vector machines." Ain Shams Engineering Journal.
- Baghban, A., S. Namvarrechi, L. T. K. Phung, M. Lee, A. Bahadori and T. Kashiwao** (2016). "Phase equilibrium modelling of natural gas hydrate formation conditions using LSSVM approach." Petroleum Science and Technology **34**(16): 1431-1438.

- Bahadori, A., A. Baghban, M. Bahadori, M. Lee, Z. Ahmad, M. Zare and E. Abdollahi** (2016). "Computational intelligent strategies to predict energy conservation benefits in excess air controlled gas-fired systems." Applied Thermal Engineering **102**: 432-446.
- Baş, D. and I. H. Boyacı** (2007). "Modeling and optimization I: Usability of response surface methodology." Journal of food engineering **78**(3): 836-845.
- Baylar, A., D. Hanbay and M. Batan** (2009). "Application of least square support vector machines in the prediction of aeration performance of plunging overfall jets from weirs." Expert Systems with Applications **36**(4): 8368-8374.
- Belghait, A., C. Si-Moussa, M. Laidi and S. Hanini** (2018). "Semi-empirical correlation of solid solute solubility in supercritical carbon dioxide: Comparative study and proposition of a novel density-based model." Comptes Rendus Chimie.
- Bovard, S., M. Abdi, M. R. K. Nikou and A. Daryasafar** (2017). "Numerical investigation of heat transfer in supercritical CO<sub>2</sub> and water turbulent flow in circular tubes." The Journal of Supercritical Fluids **119**: 88-103.
- Celso, F. L., A. Triolo, F. Triolo, J. McClain, J. Desimone, R. Heenan, H. Amenitsch and R. Triolo** (2002). "Industrial applications of the aggregation of block copolymers in supercritical CO<sub>2</sub>: a SANS study." Applied Physics A **74**(1): s1427-s1429.
- Cortes, C. and V. Vapnik** (1995). "Support-vector networks." Machine learning **20**(3): 273-297.
- Dartiguelongue, A., A. Leybros and A. s. Grandjean** (2016). "Solubility of Perfluoropentanoic Acid in Supercritical Carbon Dioxide: Measurements and Modeling." Journal of Chemical & Engineering Data **61**(11): 3902-3907.
- Eberhart, R. and J. Kennedy** (1995). A new optimizer using particle swarm theory. Micro Machine and Human Science, 1995. MHS'95., Proceedings of the Sixth International Symposium on, IEEE.
- Erkey, C.** (2000). "Supercritical carbon dioxide extraction of metals from aqueous solutions: a review." The Journal of Supercritical Fluids **17**(3): 259-287.
- Fei, C. and J. Olsen** (2011). "Prenatal exposure to perfluorinated chemicals and behavioral or coordination problems at age 7 years." Environmental health perspectives **119**(4): 573.
- Gao, W., M. Abdi-khanghah, M. Ghorogi, A. Daryasafar and M. Lavasani** (2018). "Flow reversal of laminar mixed convection for supercritical CO<sub>2</sub> flowing vertically upward in the entry region of asymmetrically heated annular channel." The Journal of Supercritical Fluids **131**: 87-98.
- Ghaziaskar, H. S., S. Afsari, M. Rezayat and H. Rastegari** (2017). "Quaternary solubility of acetic acid, diacetin and triacetin in supercritical carbon dioxide." The Journal of Supercritical Fluids **119**: 52-57.

- Ghaziaskar, H. S. and M. Kaboudvand** (2008). "Solubility of trioctylamine in supercritical carbon dioxide." The Journal of Supercritical Fluids **44**(2): 148-154.
- Ghaziaskar, H. S. and M. Nikravesh** (2003). "Solubility of hexanoic acid and butyl acetate in supercritical carbon dioxide." Fluid phase equilibria **206**(1-2): 215-221.
- Gunn, S. R.** (1998). "Support vector machines for classification and regression." ISIS technical report **14**(1): 5-16.
- Guo, H., X. Zhuang and T. Rabczuk** (2019). "A Deep Collocation Method for the Bending Analysis of Kirchhoff Plate." CMC-COMPUTERS MATERIALS & CONTINUA **59**(2): 433-456.
- Gurdial, G. S. and N. R. Foster** (1991). "Solubility of o-hydroxybenzoic acid in supercritical carbon dioxide." Industrial & Engineering Chemistry Research **30**(3): 575-580.
- Haratipour, P., A. Baghban, A. H. Mohammadi, S. H. H. Nazhad and A. Bahadori** (2017). "On the estimation of viscosities and densities of CO<sub>2</sub>-loaded MDEA, MDEA+ AMP, MDEA+ DIPA, MDEA+ MEA, and MDEA+ DEA aqueous solutions." Journal of Molecular Liquids **242**: 146-159.
- Hintzer, K., M. Juergens, G. J. Kaempf, H. Kaspar, K. H. Lochhaas, A. Streiter, O. Shyshkov, T. C. Zipplies and H. Koenigsmann** (2016). Fluoropolymer compositions and purification methods thereof, Google Patents.
- Hintzer, K., G. Löhr, A. Killich and W. Schwertfeger** (2004). Method of making an aqueous dispersion of fluoropolymers, Google Patents.
- Huang, Z., Y. C. Chiew, W.-D. Lu and S. Kawi** (2005). "Solubility of aspirin in supercritical carbon dioxide/alcohol mixtures." Fluid Phase Equilibria **237**(1): 9-15.
- Hubbard, H., Z. Guo, K. Krebs, S. Metzger, C. Mocka, R. Pope and N. Roache** (2012). "Removal of perfluorocarboxylic acids (PFCAs) from carpets treated with stain-protection products by using carpet cleaning machines." US EPA Report EPA/600/R-12/703.
- Inomata, H., Y. Honma, M. Imahori and K. Arai** (1999). "Fundamental study of de-solventing polymer solutions with supercritical CO<sub>2</sub>." Fluid phase equilibria **158**: 857-867.
- Jang, J.-S. R., C.-T. Sun and E. Mizutani** (1997). "Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence."
- Kennedy, J.** (2010). Particle swarm optimization Encyclopedia of Machine Learning (pp. 760-766), Springer.
- Khosravi, A., R. Nunes, M. Assad and L. Machado** (2018). "Comparison of artificial intelligence methods in estimation of daily global solar radiation." Journal of cleaner production **194**: 342-358.

- Knez, Z. e., D. Cör and M. a. Knez Hrnčič** (2017). "Solubility of Solids in Sub-and Supercritical Fluids: A Review 2010–2017." Journal of Chemical & Engineering Data.
- Kumoro, A. C.** (2011). "Solubility of corosolic acid in supercritical carbon dioxide and its representation using density-based correlations." Journal of Chemical & Engineering Data **56**(5): 2181-2186.
- Lin, F., D. Liu, S. Maiti Das, N. Prempeh, Y. Hua and J. Lu** (2014). "Recent progress in heavy metal extraction by supercritical CO<sub>2</sub> fluids." Industrial & Engineering Chemistry Research **53**(5): 1866-1877.
- Mehdizadeh, B. and K. Movagharnejad** (2011). "A comparative study between LS-SVM method and semi empirical equations for modeling the solubility of different solutes in supercritical carbon dioxide." Chemical Engineering Research and Design **89**(11): 2420-2427.
- Moody, C. A. and J. A. Field** (1999). "Determination of perfluorocarboxylates in groundwater impacted by fire-fighting activity." Environmental science & technology **33**(16): 2800-2806.
- Movagharnejad, K., B. Mehdizadeh, M. Banhashemi and M. S. Kordkheili** (2011). "Forecasting the differences between various commercial oil prices in the Persian Gulf region by neural network." Energy **36**(7): 3979-3984.
- Muller, K.-R., S. Mika, G. Ratsch, K. Tsuda and B. Scholkopf** (2001). "An introduction to kernel-based learning algorithms." IEEE transactions on neural networks **12**(2): 181-201.
- Munshi, P. and S. Bhaduri** (2009). "Supercritical CO<sub>2</sub>: a twenty-first century solvent for the chemical industry." Current Science (00113891) **97**(1).
- Nahar, L. and S. D. Sarker** (2012). Supercritical fluid extraction in natural products analyses. Natural Products Isolation, Springer: 43-74.
- Ohde, H., F. Hunt and C. M. Wai** (2001). "Synthesis of silver and copper nanoparticles in a water-in-supercritical-carbon dioxide microemulsion." Chemistry of materials **13**(11): 4130-4135.
- Rabczuk, T., H. Ren and X. Zhuang** (2019). "A Nonlocal Operator Method for Partial Differential Equations with Application to Electromagnetic Waveguide Problem." Computers, Materials & Continua **59** (2019), Nr. 1.
- Razavi, R., A. Sabaghmoghadam, A. Bemani, A. Baghban, K.-w. Chau and E. Salwana** (2019). "Application of ANFIS and LSSVM strategies for estimating thermal conductivity enhancement of metal and metal oxide based nanofluids." Engineering Applications of Computational Fluid Mechanics **13**(1): 560-578.
- Richter, H. P. and E. J. Dibble** (1983). Repellent coatings for optical surfaces, Google Patents.
- Rostami, A., A. Baghban and S. Shirazian** (2019). "On the Evaluation of Density of Ionic Liquids: Towards a Comparative Study." Chemical Engineering Research and Design.

- Rousseeuw, P. J. and A. M. Leroy** (2005). Robust regression and outlier detection, John Wiley & sons.
- Sahihi, M., H. S. Ghaziaskar and M. Hajebrahimi** (2010). "Solubility of maleic acid in supercritical carbon dioxide." Journal of Chemical & Engineering Data **55**(7): 2596-2599.
- Smith, M.** (1993). Neural networks for statistical modeling, Thomson Learning.
- Sovova, H., M. Zarevucka, M. Vacek and K. Stránský** (2001). "Solubility of two vegetable oils in supercritical CO<sub>2</sub>." the Journal of Supercritical fluids **20**(1): 15-28.
- Sparks, D. L., L. A. Estévez, R. Hernandez, K. Barlow and T. French** (2008). "Solubility of nonanoic (pelargonic) acid in supercritical carbon dioxide." Journal of Chemical & Engineering Data **53**(2): 407-410.
- Sparks, D. L., R. Hernandez, L. A. Estévez, N. Meyer and T. French** (2007). "Solubility of azelaic acid in supercritical carbon dioxide." Journal of Chemical & Engineering Data **52**(4): 1246-1249.
- Stassi, A., R. Bettini, A. Gazzaniga, F. Giordano and A. Schiraldi** (2000). "Assessment of solubility of ketoprofen and vanillic acid in supercritical CO<sub>2</sub> under dynamic conditions." Journal of Chemical & Engineering Data **45**(2): 161-165.
- Sunarso, J. and S. Ismadji** (2009). "Decontamination of hazardous substances from solid matrices and liquids using supercritical fluids extraction: a review." Journal of hazardous materials **161**(1): 1-20.
- Suykens, J. A. and J. Vandewalle (1999). "Least squares support vector machine classifiers." Neural processing letters **9**(3): 293-300.
- Suykens, J. A., J. Vandewalle and B. De Moor** (2001). "Optimal control by least squares support vector machines." Neural networks **14**(1): 23-35.
- Tian, G.-h., J.-s. Jin, J.-j. Guo and Z.-t. Zhang** (2007). "Mixed solubilities of 5-sulfosalicylic acid and p-aminobenzoic acid in supercritical carbon dioxide." Journal of Chemical & Engineering Data **52**(5): 1800-1802.
- Üzer, S., U. Akman and Ö. Hortaçsu** (2006). "Polymer swelling and impregnation using supercritical CO<sub>2</sub>: a model-component study towards producing controlled-release drugs." The Journal of supercritical fluids **38**(1): 119-128.
- Vapnik, V. (1998). Statistical learning theory. 1998, Wiley, New York.
- Wang, Y., S. Zhang, Z. Liu, H. Li and L. Wang** (2005). "Characterization and expression of AmphicL encoding cathepsin L proteinase from amphioxus *Branchiostoma belcheri tsingtauense*." Marine Biotechnology **7**(4): 279-286.

**Zamen, M., A. Baghban, S. M. Pourkiaei and M. H. Ahmadi** (2019). "Optimization methods using artificial intelligence algorithms to estimate thermal efficiency of PV/T system." Energy Science & Engineering.

**Zarei, F., R. Razavi, A. Baghban and A. H. Mohammadi** (2019). "Evolving generalized correlations based on Peng-Robinson equation of state for estimating dynamic viscosities of alkanes in supercritical region." Journal of Molecular Liquids **284**: 755-764.

**Zhang, X., S. Heinonen and E. Levänen** (2014). "Applications of supercritical carbon dioxide in materials processing and synthesis." Rsc Advances **4**(105): 61137-61152.

**Zhao, Z., X. Zhang, K. Zhao, P. Jiang and Y. Chen** (2017). "Numerical investigation on heat transfer and flow characteristics of supercritical nitrogen in a straight channel of printed circuit heat exchanger." Applied Thermal Engineering **126**: 717-729.