

Critical Properties and Acentric Factors of Pure Compounds Modelling Based on QSPR-SVM with Dragonfly Algorithm

M. Moussaoui,^{a,b,*} M. Laidi,^a S. Hanini,^a
A. Abdellah El Hadj,^c and M. Hentabli^d

^aLaboratory of Biomaterials and Transport Phenomena (LBMPT),
University of Médéa, Médéa, Algeria

^bUniversity of Bouira, Bouira, Algeria

^cDepartment of Chemistry, University of Blida1, Blida, Algeria

^dLaboratory Quality Control, Physico-Chemical Department, Antibiotic Sidal of Médéa, Algeria

This work is licensed under a
Creative Commons Attribution 4.0
International License



Abstract

This work aimed to model the critical pressure, temperature, volume properties, and acentric factors of 6700 pure compounds based on five relevant descriptors and two thermodynamic properties. To that end, four methods were used, namely, multi-linear regression (MLR), artificial neural networks (ANNs), support vector machines (SVMs) using sequential minimal optimisation (SMO), and hybrid SVM with Dragonfly optimisation algorithm (SVM-DA) to model each property. The results suggested that hybrid SVM-DA had better prediction performance compared to the other models in terms of average absolute relative deviation (AARD%) of {0.7551, 1.962, 1.929, and 2.173} and R^2 of {0.9699, 0.9673, 0.9856, and 0.9766} for critical temperature, critical pressure, critical volume, and acentric factor, respectively. The developed models can be used to estimate the property of newly designed compounds only from their molecular structure.

Keywords

Support vector machine, critical properties, Dragonfly optimisation algorithm, quantitative structure-property relationship

1 Introduction

Physical properties of chemical compounds, such as critical properties and acentric factor are important for scientists in designing and modelling chemical processes¹ such as supercritical extraction and their importance in the application of different equation of states. Experimental computing of these properties is costly, time-consuming, and sometimes even impossible, given that some larger compounds or higher molecular weight may chemically degrade before they reach critical conditions.¹ Therefore, different predicting methods have been published in the literature to compute these critical properties when experimental values are not available. To date, two approaches have been proposed to estimate critical properties, group contribution, and quantitative structure-property relationship (QSPR) models. Since group contribution based models have some limitations,² another approach based on QSPR was used to estimate the physical properties solely from the molecular descriptors, which are computed by applying certain mathematical algorithms on the molecular structure of components.²⁻⁴ Several approaches have been employed to estimate these properties, namely, group contribution from molecular structures,^{5,6} multi-linear regression based on QSPR approaches from descriptors.¹ In addition to different computational techniques like artificial neural networks (ANNs),^{7,8} adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM) have been successfully used to estimate these properties of different components from descriptors.^{1,9-12}

A well-known intelligent learning method, such as SVM, is used as a powerful predictive tool to fit the non-linear behaviour of any system.¹³ This approach has great modelling capability when large datasets are available and the relationships of parameters are complicated.¹⁴ In addition, SVM approach was originally developed by Cortes and Vapnik¹⁵ and identified as a low parameter technique with a rapid calculation process to find the best tuning parameters and robust performance for overcoming the overfitting problem compared with traditional intelligent learning algorithms.

Therefore, the novelty of this work is in the use of a hybrid approach between SVMs and Dragonfly optimisation algorithm (DA) to model the critical properties and acentric factors of pure compounds based on a large and representative data set.

2 Materials and methods

In this work, a dataset of 6700 chemical compounds was collected from previously published papers and books in the literature.^{1,16} Descriptors were computed using *alvaDesc* software based on the simplified molecular-input line-entry system (SMILES) of each compound, which computes 5305 types of descriptors, (*i.e.*, 0D, 1D, 2D, and 3D descriptors).¹⁷ The relevant descriptors were pre-selected in two steps. The first selection was performed using *alvaDesc* software by reducing all descriptors as follows:

* Corresponding author: Mohammed Moussaoui, PhD
Email: moussaoui24@gmail.com, m.moussaoui@univ-bouira.dz

- Constant and near constant descriptor values were removed;
- All missing values or at least one missing value were removed;
- Any descriptor with a relative standard deviation < 0.001 was removed;
- Descriptors with pair correlation larger or equal to ≥ 0.95 were removed.

The second step consisted of using stepwise MLR as final screening to select only the relevant descriptors. Five descriptors remained after the final screening, as follows: $\{S_p$: sum of atomic polarizabilities (scaled on carbon atom), M_p : mean atomic polarizability (scaled on carbon atom), GD : graph density, $H\%$ percentage of H atoms, and $C\%$: percentage of C atoms}. The number of inputs for each property was 5 descriptors, boiling temperatures (T_b), and molecular weights (M_w).

The performances of SVM method for regression depend mainly on several parameters and the kernel function. Since the relationship among selected molecular descriptors and the selected properties is non-linear and for regression tasks, a commonly used kernel function is the Gaussian radial basis function (RBF) due to its good general performance – this method is well explained in the following paper.¹⁸

2.1 Description of SVM and DA method

SVMs are a supervised learning method presented for the first time in 1992 by B. E. Boser et al.¹⁹ A few years later, they were introduced by C. Cortes and V. Vapnik¹⁵ for classification cases. In 1995 they were expanded and adjusted to regression problems by V. Vapnik²⁰ to map the non-linear relationship between inputs and outputs, and construct the regression estimation function in high-dimensional feature space $\Phi(x)$, map back to the original space and define the suitable kernel function $K(x_i, y_i)$, where x_i is the input value, y_i is the output value, and $\{i = 1, \dots, m\}$ m is the size of the learning dataset $D = (x_i, y_i)$.²¹ The regression function is designed by $y = W^T \cdot \Phi(x) + b$, where W is the weight vector and b is the bias.²² The major advantages of SVM model are the absence of local minima during the training stage, it deals with non-linear regression, sparse, employs kernel functions.²³ More details about the SVM model optimisation equation and the kernel functions can be found in ref.²⁴⁻²⁶ A novel metaheuristic optimisation algorithm named Dragonfly (DA) was developed by S. Mirjalili²⁷ based on the behaviour of dragonflies.²⁷ This algorithm was coupled with SVM method to tune its hyper-parameters.²⁸

3 Results and discussion

3.1 Critical temperature

The predicted critical temperature values using the developed MLR model was compared to the experimental values using two metrics, average absolute relative deviation (AARD%) and determination coefficient (R). Fig. 1(a)

shows the performance of the MLR model in terms of $\{AARD\% = 3.376\%$ and $R^2 = 0.9104\}$. In this figure, the red solid line indicates the exact fit between the experimental and predicted properties, whereas the black circles demonstrate the training points, and the blue squares show the test points correlated for all properties by the proposed model vs experimental data. The closer the points to the solid line, the more accurate correlated properties data. The equation of this model is written as follows:

$$T_c = -18.93 - 0.0977 M_w + 1.47 T_b + 3.60 S_p + 55.56 M_p + 31.90 GD + 0.28 H\% + 0.57 C\% \quad (1)$$

Secondly, the parameters of the best ANN model were determined based on a trial-and-error method. Results showed that the best multi-layer perception feed-forward back-propagation neural network trained with Levenberg–Marquardt algorithm (MLP-ANN) was obtained with the architecture of $\{7, 15, 1\}$ neurons in the first layer, hidden layer, and output layer, respectively. Tangent hyperbolic (f) and linear (g) transfer functions were used in the hidden layer and the output layer, respectively. The regression curve is presented in Fig. 1(b), where the performance of the ANN model in terms of $\{AARD\% = 2.639\%$ and $R^2 = 0.9450\}$ and its expression can be written as follows:

$$T_c = g \left(\sum_{j=1}^{15} W_{1,j}^S f \left(\sum_{i=1}^7 W_{j,i}^I X_i + b_{hj} \right) + b_{o1} \right) \quad (2)$$

where X_i is the input vector, $W_{j,i}^I, W_{1,j}^S$ are input-hidden layer and hidden-output layer weights matrix, and b_{hj}, b_{o1} are biases of hidden and output layers, respectively.

Another model was developed based on the SVM approach, the best model was based on the optimisation of its parameters, namely, capacity (Box Constraint) (C), the kernel width parameter (Kernel scale) (γ), the quantity of support vectors (QSV) and the kernel function. The accurate model was found with RBF kernel function and $\{C = 10, \gamma = 1, \text{ and } QSV = 4008\}$ for the critical temperature. This model was found with $\{AARD\% = 1.224\%$ and $R^2 = 0.9375\}$ as reported in Fig. 1(c).

The last model was based on the use of SVM hybrid with recent optimisation algorithm named Dragonfly, which was inserted in the *MATLAB* software. The graphical comparison between the predicted and the experimental critical temperatures was conducted using the *postreg MATLAB* function. Based on the obtained results presented in Fig. 1(d), the proposed SVM-DA model can highly model the critical temperature in terms of $\{AARD\% = 0.7551\%$ and $R^2 = 0.9699\}$ in comparison to the other developed models. The obtained SVM-DA was found with RBF kernel function and $\{C = 10, \gamma = 0.5, \text{ and } QSV = 3293\}$.

3.2 Critical pressure

Results showed that the MLR model was able to predict the critical pressure values with $\{AARD\% = 10.47\%$ and $R^2 = 0.7821\}$. The critical pressure can be written in func-

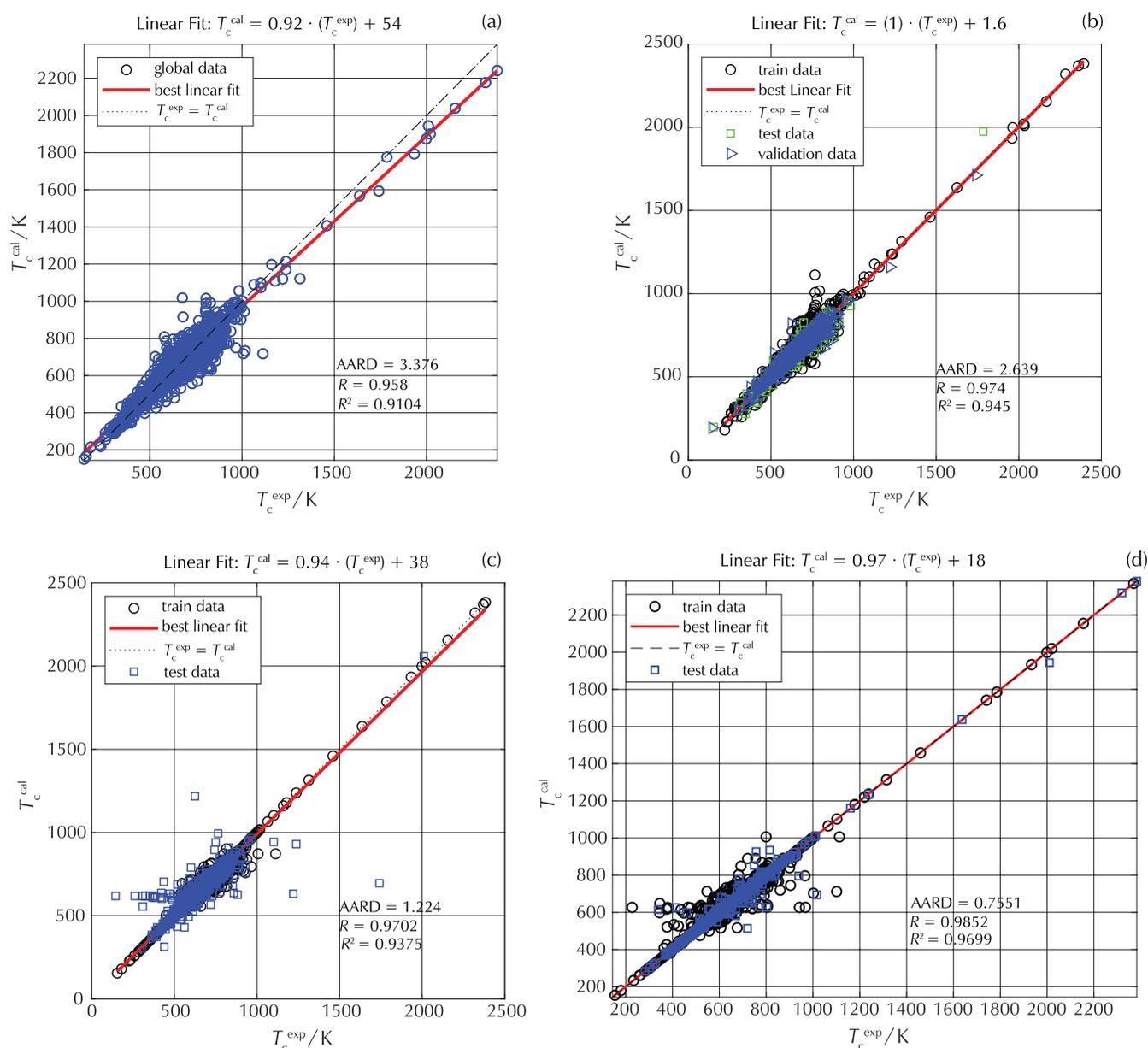


Fig. 1 – Comparison between experimental and predicted critical temperature of the entire database: (a) MLR, (b) ANN, (c) SVM, and (d) SVM-DA

tion of the selected inputs via the best MLR model as follows:

$$p_c = 18.96 - 0.0751 M_w + 0.0734 T_b - 1.13 S_p + 10.97 M_p + 47.62 GD - 0.124 H\% - 0.1548 C\% \quad (3)$$

For the best ANN model, results showed that the best structure of the best model was $\{7, 15, 1\}$ neurons in the first layer, hidden layer, and output layer, respectively. The regression comparison is depicted in Fig. 2(b), where the performance of the ANN model was found as $\{\text{AARD}\% = 5.897\%$ and $R^2 = 0.911\}$.

The best SVM model was obtained with the RBF kernel function and $\{C = 10, \gamma = 1, \text{ and } QSV = 4990\}$ for the critical pressure. The accuracy of this model was $\{\text{AARD}\% = 2.575\%$ and $R^2 = 0.9491\}$, and the regression plot is presented in Fig. 2(c).

The graphical comparison presented in Fig. 2(d) between the SVM-DA predicted and the experimental values of the critical pressure and the calculated statistical parameters shows the performance of this model in terms of $\{\text{AARD}\% = 1.962\%$ and $R^2 = 0.9673\}$. This model was found with a RBF kernel function and $\{C = 10, \gamma = 0.5, \text{ and } QSV = 4785\}$, and showed high capability of predicting critical pressure compared to other tested models in this study.

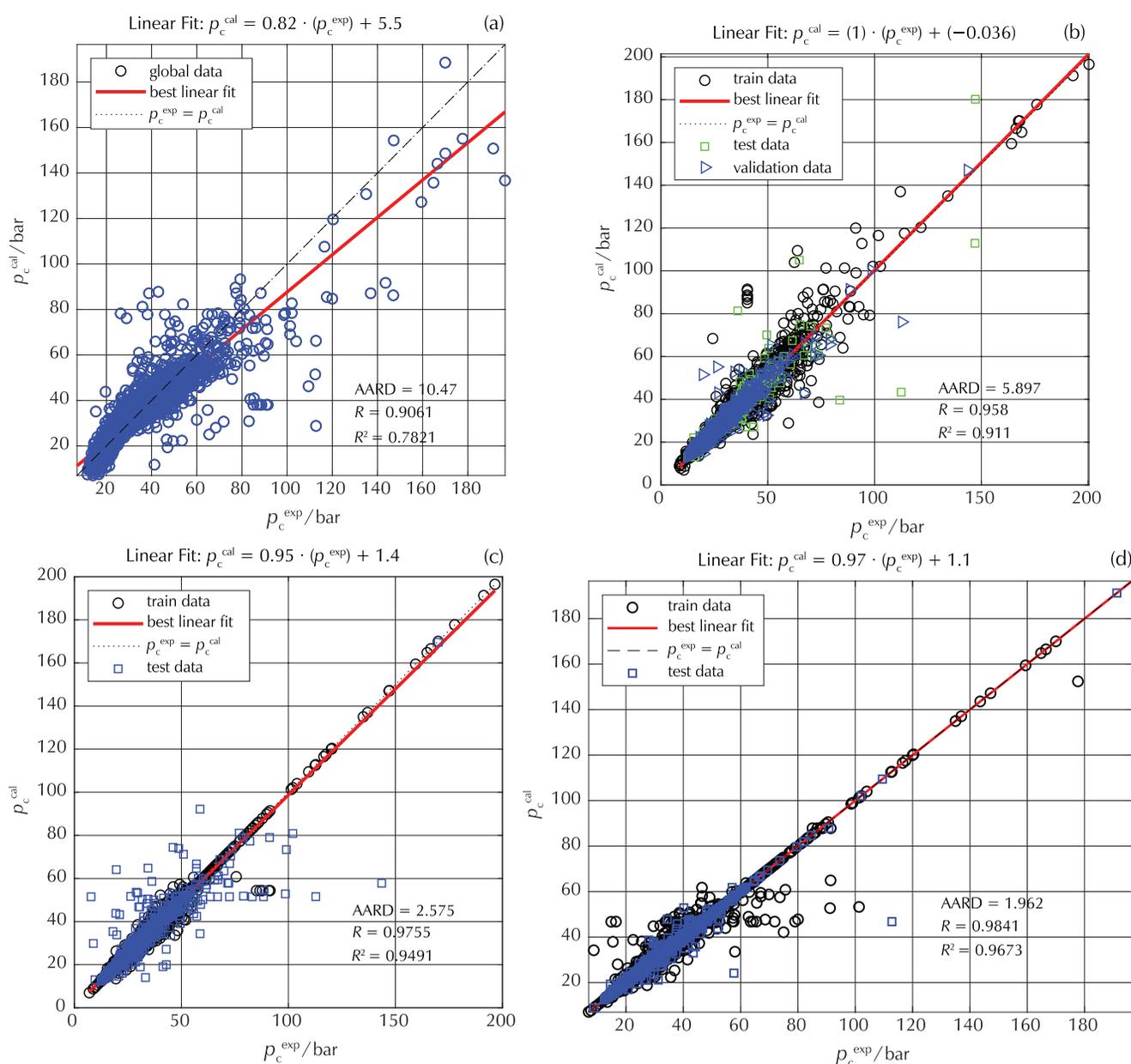


Fig. 2 – Comparison between experimental and predicted critical pressure of the entire database: (a) MLR, (b) ANN, (c) SVM, and (d) SVM-DA

3.3 Critical volume

Fig. 3(a) shows the performance of the best MLR model in terms of {AARD% = 6.627 % and $R^2 = 0.9339$ }. The equation of this model is written as follows:

$$V_c = 127.14 + 0.4065 M_w + 0.1988 T_b + 28.169 S_p - 129.90 M_p - 11.84 GD - 1.1247 H\% - 1.3638 C\% \quad (4)$$

Fig. 3(b) illustrates the regression plot and shows the accuracy of the best ANN model with {AARD% = 4.534 % and $R^2 = 0.971$ }. This model was found with the architecture of {7, 15, 1} neurons in the first layer, hidden layer, and output layer, respectively. Tangent hyperbolic (f) and linear (g) transfer function was used in the hidden layer and the output layer, respectively.

Fig. 3(c) illustrates the regression plot and shows the accuracy of the best SVM model with {AARD% = 2.338 % and $R^2 = 0.9792$ }. The accurate model was found with RBF kernel function and { $C = 10$, $\gamma = 0.5$, and $QSV = 4817$ } for the critical volume.

Fig. 3(d) shows the regression plot between experimental and predicted critical volume using SVM-DA model. Results show that SVM-DA can highly model the critical volume with {AARD% = 1.929 % and $R^2 = 0.9856$ } in comparison to the other developed models. The obtained SVM-DA was found with RBF kernel function and { $C = 10$, $\gamma = 0.5$, and $QSV = 4795$ }.

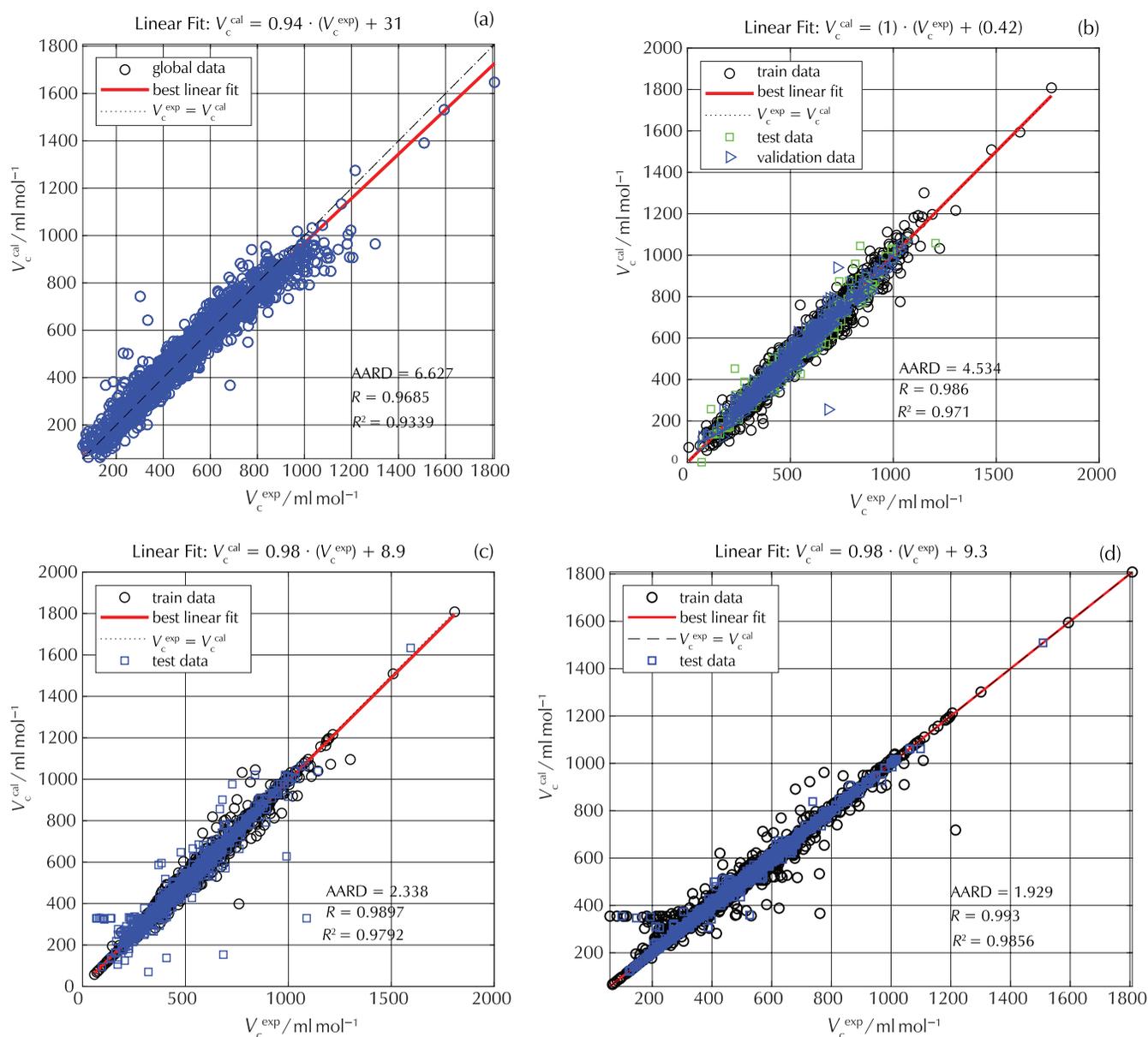


Fig. 3 – Comparison between experimental and predicted critical volume of the entire database: (a) MLR, (b) ANN, (c) SVM, and (d) SVM-DA

3.4 Acentric factor

The acentric factor (w) was modelled using the best MLR model with the performance of {AARD% = 23.76 % and $R^2 = 0.232$ }, as presented in Fig. 4(a). The equation of this model is written as follows:

$$w = 0.3369 - 6.86 \cdot 10^{-5} M_w + 1.25 \cdot 10^{-3} T_b + 9.858 \cdot 10^{-3} S_p - 0.333 M_p - 0.246 GD - 1.448 \cdot 10^{-3} H\% - 5.069 \cdot 10^{-3} C\% \quad (5)$$

The acentric factor was modelled using the best ANN model with the performance of {AARD% = 9.733 % and $R^2 = 0.890$ }, as presented in Fig. 4(b). This model was optimised with the structure of {7, 20, 1} neurons in the first layer, hidden layer, and output layer, respectively. Tangent hyperbolic (f) and linear (g) transfer functions were used in

the hidden layer and the output layer, respectively.

The acentric factor was modelled using the best SVM model with the performance of {AARD% = 6.140 % and $R^2 = 0.8725$ }, as presented in Fig. 4(c). The accurate model was found with RBF kernel function and { $C = 10$, $\gamma = 1$, and QSV = 4349}.

The acentric factor was modelled using the best SVM-DA model with the performance of {AARD% = 2.173 % and $R^2 = 0.9766$ }, as presented in Fig. 4(a). The obtained SVM-DA was found with RBF kernel function and { $C = 100$, $\gamma = 0.5$, and QSV = 3032}.

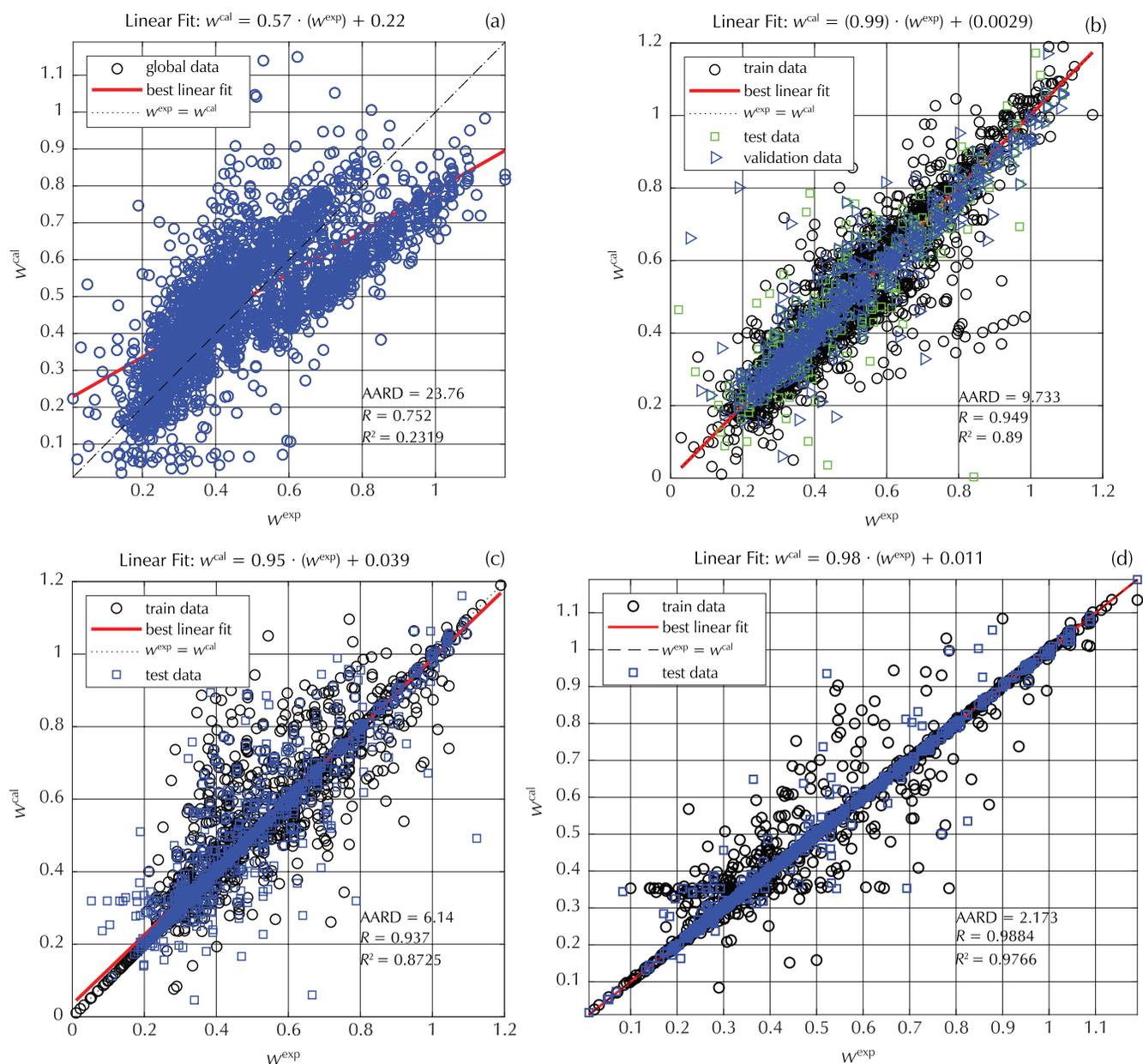


Fig. 4 – Comparison between experimental and predicted acentric factor of the entire database: (a) MLR, (b) ANN, (c) SVM, and (d) SVM-DA

All values of the critical properties predicted through each model and corresponding experimental values are presented as supplementary data.

As may be seen in Fig. 5, the SVM-DA model was able to predict the selected parameters with high accuracy in comparison to the other models. Therefore, a sensitivity analysis, the applicability domain, and the statistical evaluation and validation section were performed based on the obtained SVM-DA parameters.

3.5 Applicability domain of the developed models

In this study, the applicability domain is presented using William's plot. This plot illustrates the standardised residu-

als against the leverage values (h). The leverage of the i^{th} (h_i) component is defined as follows:

$$h_i = X_i (X^T X)^{-1} X_i^T \quad i = 1, 2, 3 \dots n \quad (6)$$

where X is the descriptor matrix of the training set, X_i is the descriptor vector for the desired compound, the superscript means transpose, and n is the number of molecular structures in the training set. In this plot, the critical leverage (h^*) was computed according to the following equation:

$$h^* = 3(P+1)/n \quad (7)$$

where P is the number of descriptors in the model. The standardised residuals δ were calculated by:

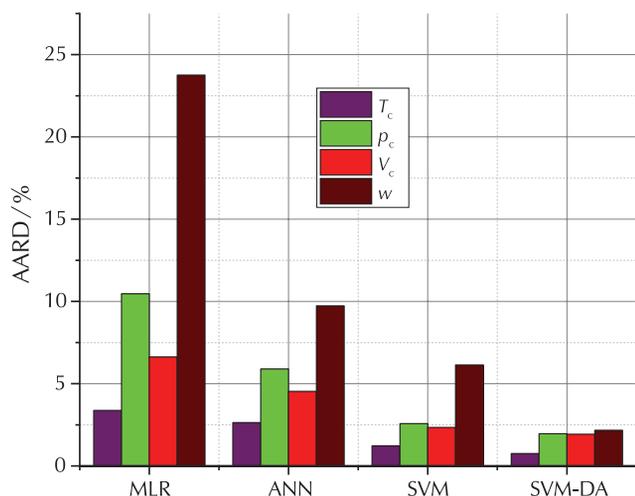
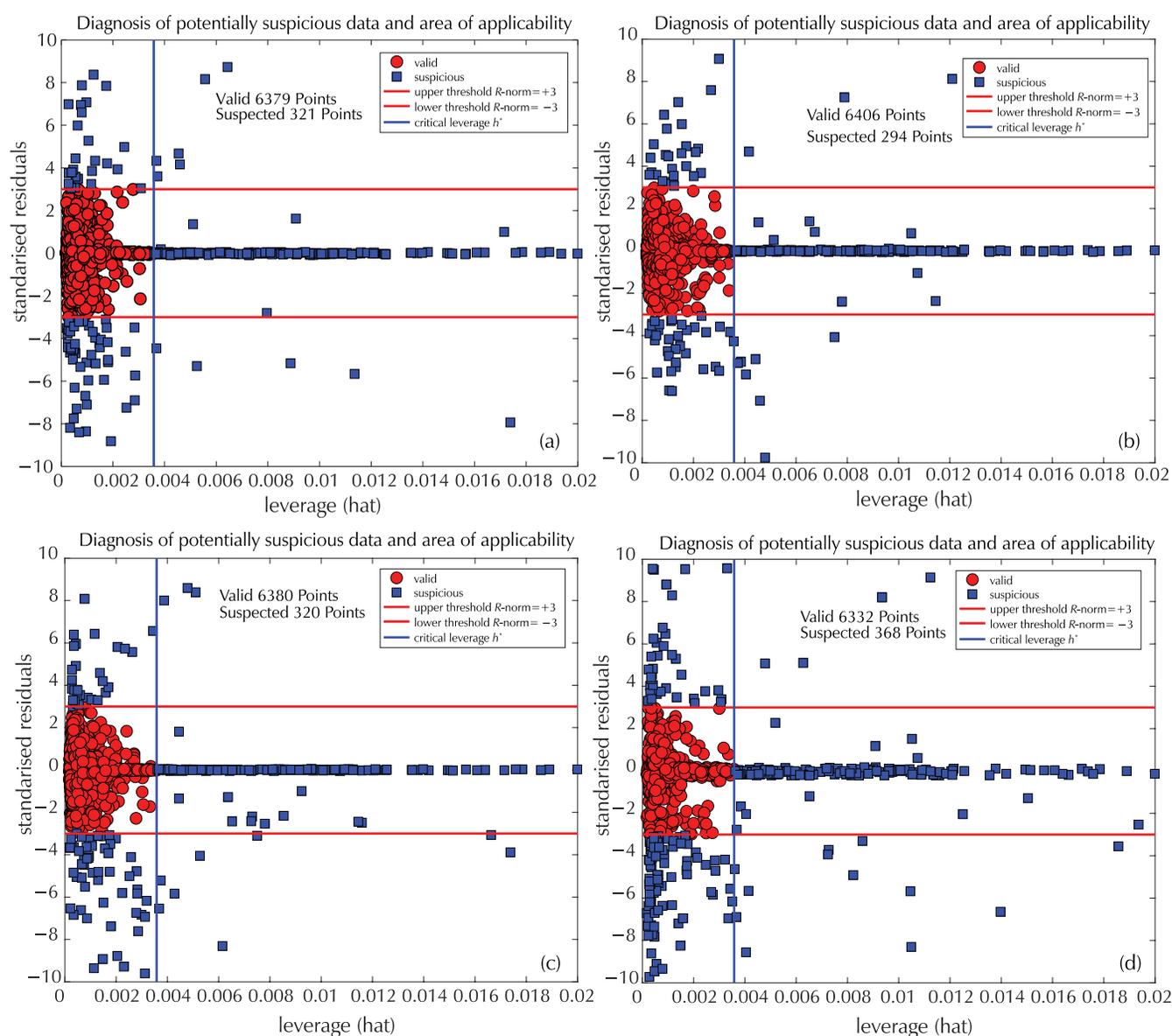


Fig. 5 – Model prediction performance of four parameters

$$\delta = \frac{y_i - \hat{y}_i}{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - P - 1)}} \quad (8)$$

where y_i and \hat{y}_i are the experimental value and calculated value for the i^{th} compound, respectively.

The outliers are data points when the values of standardised residual are in the range of ± 3 standard deviations units and leverage value should be higher than h^* . As may be observed in Fig. 6, the outliers represent only about 4.791, 4.382, 4.772, and 5.492 % of total compounds in the data set for T_c , p_c , V_c and w , respectively. The developed models were capable of predicting the majority of experimental data with values of standardised residual lower than ± 3 . A few points are out of the application range with larger values of h or standardised residuals. However, outliers do not mean that the prediction capability of the model is not good, the AARD% and R are a judgement of the predictive ability for SVM-DA model.

Fig. 6 – William's plot for each related property (a) T_c , (b) p_c , (c) V_c , and (d) w

3.6 Statistical evaluation and validation of the developed models

Fig. 7 illustrates the values of absolute relative deviation ARD% of the modelled properties. Results proved that these models made a good prediction by yielding a very low error of almost all the dataset, and confirmed that all the selected descriptors for each model were not chosen by chance.

3.7 Global sensitivity analysis

The objective of the global sensitivity analysis is to assess the effect of inputs on each output using the cosine amplitude method (CAM).^{29–31} This method is based on the calculation of the relevance factor (r_{ij}) which measures the strength relationships between the output x_j and input x_i ,

parameters. This factor can be represented by the following expression:²⁶

$$r_{ij} = \frac{\sum_{k=1}^m X_{ik} X_{jk}}{\sqrt{\sum_{k=1}^m X_{ij}^2 \sum_{k=1}^m X_{jk}^2}} \quad (9)$$

The used data pairs build a data array X , which are defined as $X = \{X_1, X_2, X_3, \dots, X_m\}$ where X_i is a vector of lengths m , expressed as: $X_i = \{x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}\}$. i and k are the dimensions of the input matrix ($i = 1$ to 6700, $k = 1$ to m), j represents the dimensions of the output vector ($j = 1$ to 6700), and m is the number of input parameters ($m = 7$). Fig. 8 shows the importance of the selected inputs on each output. It is clearly shown that all parameters have approximately the same effect on each output.

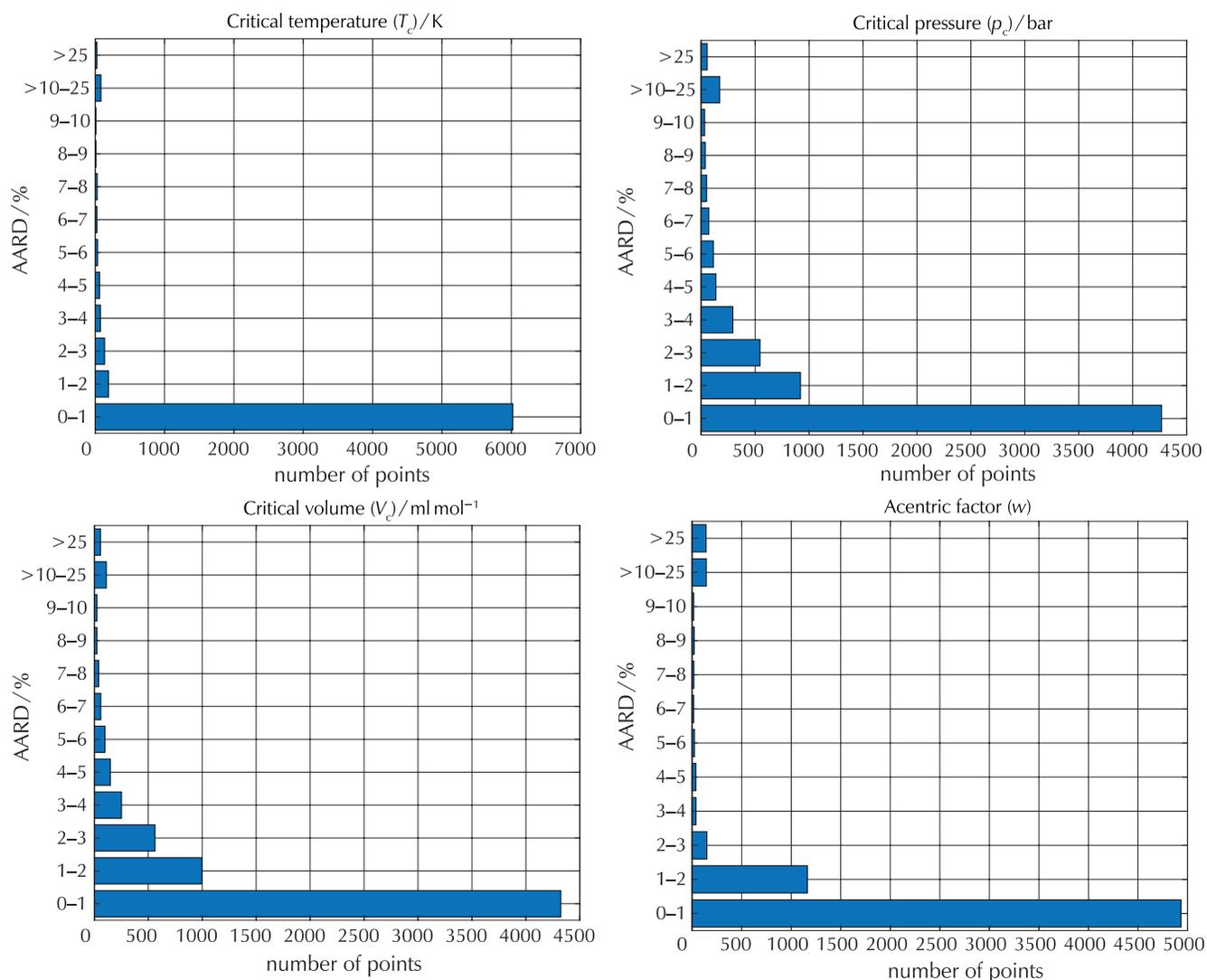


Fig. 7 – Absolute relative deviation (ARD%) of the proposed models for the prediction of T_c , p_c , V_c , and w of about 6700 data points, $ARD\% = \left| \frac{y_i^{\text{exp}} - y_i^{\text{pred}}}{y_i^{\text{exp}}} \right| \times 100$

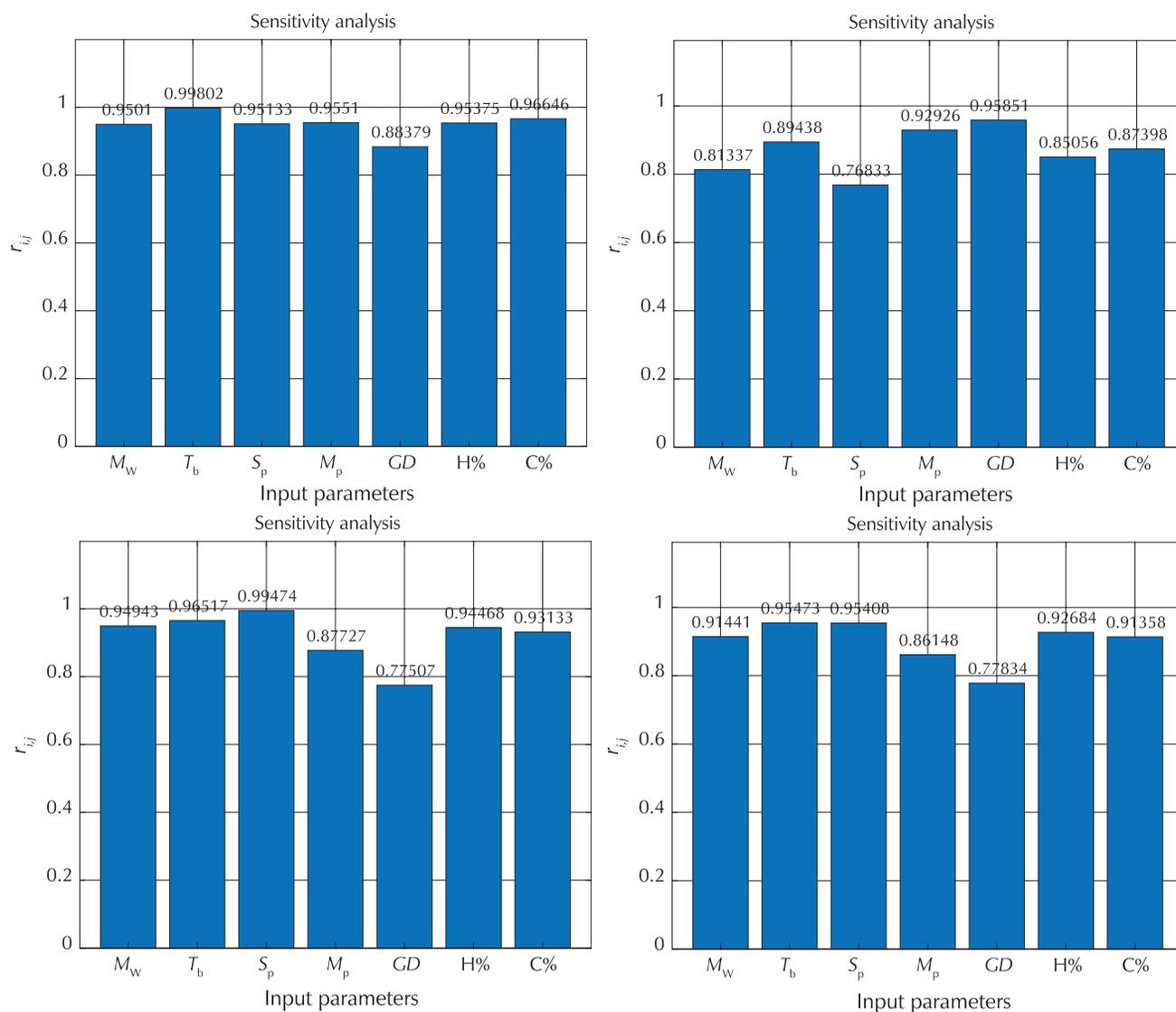


Fig. 8 – Importance of inputs on each output

Fig. 9 shows another comparison of a sample of compounds during the test stage between the experimental literature properties (shown as white face markers) and the SVR-DA predicted properties (shown as full face markers). This figure shows an excellent agreement between pure compounds properties and those experimental.

The best SVM-DA model was implemented in friendly graphical user interface designed by *MATLAB* software, and presented in Fig. 10. This interface can compute the desired outputs by knowing only the selected inputs for 6700 systems.

4 Conclusion

The goal of this study was to test the applicability of four intelligence computational methods in modelling the critical properties and acentric factors of 6700 pure compounds.

Four models were designed, namely, MLR, ANN, SVM, and SVM-DA. The experimental dataset of the properties were collected from the literature, while the input matrix consisted of a mixture of 05 descriptors and 02 thermodynamic properties.

All models were assessed statistically in terms of AARD% and R^2 . The obtained results showed that the SVM-DA was found with a very low deviation (AARD%) of {0.7551, 1.962, 1.929, and 2.173} and high determination coefficient R^2 of {0.9699, 0.9673, 0.9856, and 0.9766} for the four properties. It can be concluded that the SVM-DA model is capable of establishing a satisfactory nonlinear relationship between the molecular descriptors and the critical properties of pure compounds with high accuracy. Based on the SVM-DA model, a sensitivity analysis investigation was performed to assess the effects of inputs on each output. The results showed that all inputs had a strong effect on the output, and their choice was based on

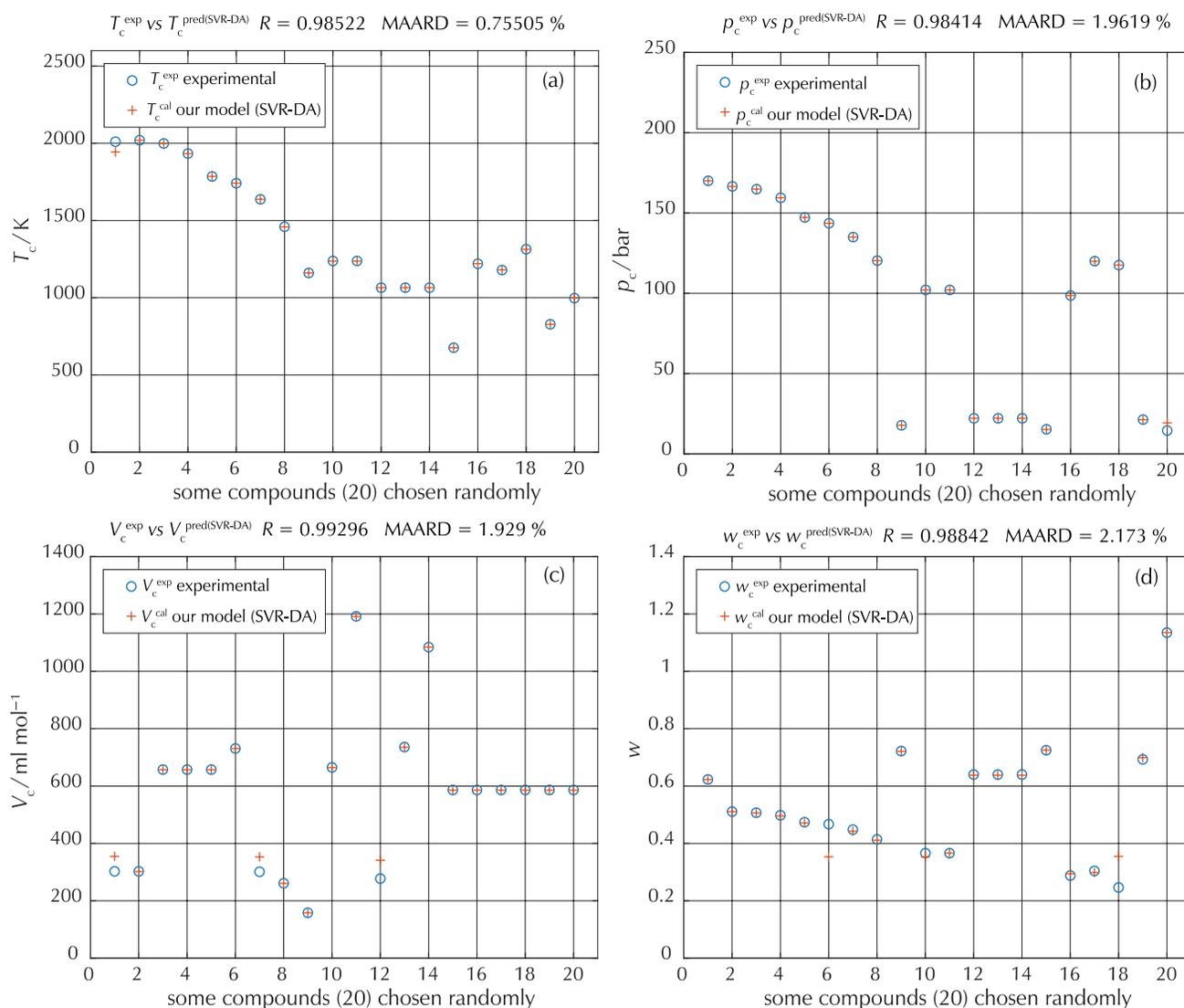


Fig. 9 – Experimental vs predicted by SVR-DA properties of some compounds during the test stage

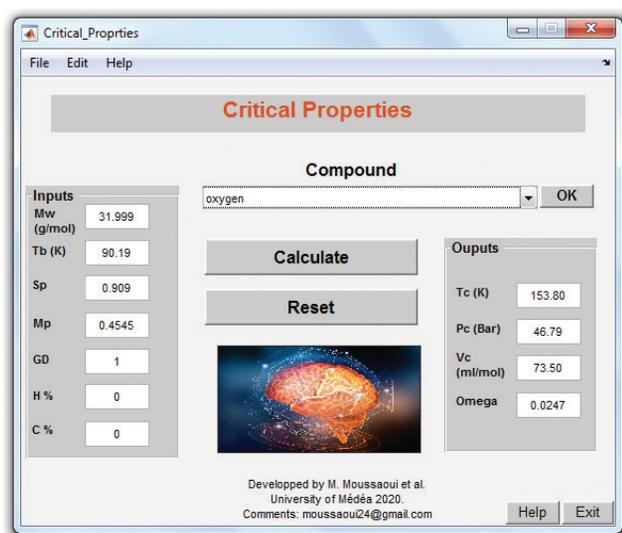


Fig. 10 – Graphical user interface

their contribution on the studied phenomenology. A separate section of this study is the applicability domain of the model against the data set used, and the results show that the SVM-DA model can be applied to fit almost 95 % of the used data, which suggests the capability of the model when modelling the studied properties in comparison to the other models. Finally, to facilitate the use of this model, a friendly graphical user interface was designed using *MATLAB* software without knowing the details about the phenomenology or the used software.

ACKNOWLEDGEMENTS

The authors wish to thank the Laboratory of Biomaterials and Transport Phenomena, University of Yahia Fares, Médéa, Algeria. The authors would also like to thank the anonymous reviewers for their constructive comments, which helped to improve the quality and presentation of this paper.

References

Literatura

1. F. Gharagheizi, A. Eslamimanesh, A. H. Mohammadi, D. Richon, Determination of critical properties and acentric factors of pure compounds using the artificial neural network group contribution algorithm, *J. Chem. Eng. Data* **56** (2011) 2460–2476, doi: <https://doi.org/10.1021/jc200019g>.
2. D. Aboali, M. A. Sobati, Novel method for prediction of normal boiling point and enthalpy of vaporization at normal boiling point of pure refrigerants: A QSPR approach, *Int. J. Refrig.* **40** (2014) 282–293, doi: <https://doi.org/https://doi.org/10.1016/j.ijrefrig.2013.12.007>.
3. B. Hemmateenejad, P. Ilani-kashkouli, Quantitative structure-property relationship study to predict speed of sound in diverse organic solvents from solvent structural information, *Ind. Eng. Chem. Res.* **51** (2012) 14884–14891, doi: <https://doi.org/10.1021/ie3016297>.
4. A. Khajeh, H. Modarress, QSPR prediction of surface tension of refrigerants from their molecular structures, *Int. J. Refrig.* **35** (2012) 150–159, doi: <https://doi.org/10.1016/j.ijrefrig.2011.08.007>.
5. A. R. Katritzky, M. Kuanar, S. Slavov, C. D. Hall, M. Karelson, I. Kahn, D. A. Dobchev, Quantitative correlation of physical and chemical properties with chemical structure: utility for prediction, *Chem. Rev.* **110** (2010) 5714–5789, doi: <https://doi.org/https://doi.org/10.1021/cr900238d>.
6. J. Marrero, R. Gani, Group-contribution based estimation of pure component properties, *Fluid Phase Equilib.* **183–184** (2001) 183–208, doi: [https://doi.org/10.1016/S0378-3812\(01\)00431-9](https://doi.org/10.1016/S0378-3812(01)00431-9).
7. M. Laidi, S. Hanini, A. Rezrazi, M. R. Yaiche, A. A. El Hadj, F. Chellali, Supervised artificial neural network-based method for conversion of solar radiation data (case study: Algeria), *Theoret. Appl. Climatol.* **128** (2017) 439–451, doi: <https://doi.org/10.1007/s00704-015-1720-7>.
8. M. Laidi, S. Hanini, Optimal solar COP prediction of a solar-assisted adsorption refrigeration system working with activated carbon/methanol as working pairs using direct and inverse artificial neural network, *Int. J. Refrig.* **36** (2013) 247–257, doi: <https://doi.org/10.1016/j.ijrefrig.2012.09.016>.
9. M. A. S. Zahra Ghomishah, Ali Ebrahimpoor Gorji, Prediction of critical properties of sulfur-containing compounds: New QSPR models, *J. Mol. Graph. Model.* (2019) 135577, doi: <https://doi.org/10.1016/j.scitotenv.2019.135577>.
10. H. Maouz, L. Khaouane, S. Hanini, Y. Ammi, QSPR studies of carbonyl, hydroxyl, polyene indices, and viscosity average molecular weight of polymers under photostabilization using ANN and MLR approaches, *Kem. Ind.* **69** (2020), doi: <https://doi.org/10.15255/KUI.2019.022>.
11. S. Keskes, S. Hanini, M. Hentabli, M. Laidi, Artificial intelligence and mathematical modelling of the drying kinetics of pharmaceutical powders, *Kem. Ind.* **69** (2020) 137–152, doi: <https://doi.org/10.15255/KUI.2019.038>.
12. N. Melzi, L. Khaouane, Y. Ammi, S. Hanini, M. Laidi, H. Zentou, Comparative study of predicting the molecular diffusion coefficient for polar and non-polar binary gas using neural networks and multiple linear regressions, *Kem. Ind.* **68** (2019) 573–582, doi: <https://doi.org/10.15255/KUI.2019.010>.
13. A. Baghban, J. Sasanipour, S. Habibzadeh, Z. Zhang, Sulfur dioxide solubility prediction in ionic liquids by a group contribution – LSSVM model, *Chem. Eng. Res. Des.* **142** (2019) 44–52, doi: <https://doi.org/10.1016/j.cherd.2018.11.026>.
14. A. Baghban, M. N. Kardani, S. Habibzadeh, Prediction viscosity of ionic liquids using a hybrid LSSVM and group contribution method, *J. Mol. Liq.* **236** (2017) 452–464, doi: <https://doi.org/10.1016/j.molliq.2017.04.019>.
15. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.* **20** (1995) 273–297, doi: <https://doi.org/10.1007/BF00994018>.
16. C. L. Yaws, *The Yaws handbook of physical properties for hydrocarbons and chemicals: physical properties for more than 54,000 organic and inorganic chemical compounds, Coverage for C1 to C100 Organics and Ac to Zr Inorganics*, Gulf Professional Publishing, 2015.
17. A. Mauri, *alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints*, in *Ecotoxicological QSARs*, Springer, 2020, pp. 801–820.
18. S. Dong, Y. Zhang, Z. He, N. Deng, X. Yu, S. Yao, Investigation of support vector machine and back propagation artificial neural network for performance prediction of the organic rankine cycle system, *Energy* **144** (2018) 851–864, doi: <https://doi.org/10.1016/j.energy.2017.12.094>.
19. B. E. Boser, I. M. Guyon, V. N. Vapnik, A training algorithm for optimal margin classifiers, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 1992, pp. 144–152, doi: <https://doi.org/10.1145/130385.130401>.
20. V. Vapnik, *The nature of statistical learning theory*, Springer Science & Business Media, 2013.
21. S. Abdollahi, H. R. Pourghasemi, G. A. Ghanbarian, R. Safaeian, Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions, *B. Eng. Geol. Environ.* **78** (2019) 4017–4034, doi: <https://doi.org/10.1007/s10064-018-1403-6>.
22. K. Yan, C. Shi, Prediction of elastic modulus of normal and high strength concrete by support vector machine, *Constr. Build. Mater.* **24** (2010) 1479–1485, doi: <https://doi.org/10.1016/j.conbuildmat.2010.01.006>.
23. J. Tinoco, A. G. Correia, P. Cortez, Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns, *Comput. Geotech.* **55** (2014) 132–140, doi: <https://doi.org/10.1016/j.compgeo.2013.08.010>.
24. H. Yan, J. Zhang, S. S. Rahman, N. Zhou, Y. Suo, Predicting permeability changes with injecting CO₂ in coal seams during CO₂ geological sequestration: A comparative study among six SVM-based hybrid models, *Sci. Total Environ.* **705** (2020) 135941, doi: <https://doi.org/doi.org/10.1016/j.scitotenv.2019.135941>.
25. H. Benimam, C. S. Moussa, M. Hentabli, S. Hanini, M. Laidi, Dragonfly-Support vector machine for regression modeling of the activity coefficient at infinite dilution of solutes in imidazolium ionic liquids using σ -profile descriptors, *J. Chem. Eng. Data* **65** (2020) 3161–3172, doi: <https://doi.org/10.1021/acs.jced.0c00168>.
26. M. Moussaoui, M. Laidi, S. Hanini, M. Hentabli, Artificial neural network and support vector regression applied in quantitative structure-property relationship modelling of solubility of solid solutes in supercritical CO₂, *Kem. Ind.* **69** (2020) 611–630, doi: <https://doi.org/10.15255/KUI.2020.004>.
27. S. Mirjalili, Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems, *Neural Comput. and Appl.* **27** (2016) 1053–1073, doi: <https://doi.org/10.1007/s00521-015-1920-1>.
28. A. Ibrir, Y. Kerchich, N. Hadidi, H. Merabet, M. Hentabli, Prediction of the concentrations of PM₁, PM_{2.5}, PM₄, and PM₁₀ by using the hybrid dragonfly-SVM algorithm, *Air Qual. Atmos. Health* (2020) 1–11, doi: <https://doi.org/10.1007/s11869-020-00936-1>.

29. E. Uzlu, A. Akpınar, H. T. Öztürk, S. Nacar, M. Kankal, Estimates of hydroelectric generation using neural networks with the artificial bee colony algorithm for Turkey, *Energy* **69** (2014) 638–647, doi: <https://doi.org/10.1016/j.energy.2014.03.059>.
30. H. Ahmadi, M. Rodehutscord, Application of artificial neural network and support vector machines in predicting metabolizable energy in compound feeds for pigs, *Front. Nutr.* **4** (2017) 1–8, doi: <https://doi.org/10.3389/fnut.2017.00027>.
31. S. Andrea, T. Stefano, C. Francesca, R. Marco, *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*, John Wiley & Sons, Ltd, 2004.

SAŽETAK

Kritična svojstva i acentrični čimbenici modeliranja čistih spojeva primjenom modela QSPR-SVM i algoritma Dragonfly

Mohammed Moussaoui,^{a,b,*} Maamar Laidi,^a Salah Hanini,^a
Abdallah Abdellah El Hadj^c i Mohamed Hentabli^d

Cilj ovog rada bio je modeliranje kritičnog tlaka, temperature, volumnih svojstava i acentričnih čimbenika 6700 čistih spojeva na temelju pet relevantnih deskriptora i dva termodinamička svojstva. U tu svrhu primijenjene su četiri metode: višestruka linearna regresija (MLR), umjetna neuronska mreža (ANN), metoda potpornih vektora (SVM) i algoritam optimizacije Dragonfly (SVM-DA), koji se za modeliranje svakog svojstva koriste sekvencijalnom minimalnom optimizacijom (SMO) i hibridnim SVM-om. Rezultati su pokazali da hibridni SVM-DA daje bolje predviđanje u odnosu na ostale modele u smislu postotka prosječnog apsolutnog relativnog odstupanja (AARD%) od {0,7551, 1,962, 1,929 i 2,173} i R^2 od {0,9699, 0,9673, 0,9856, i 0,9766} za kritičnu temperaturu, kritični tlak, kritični volumen i acentrični faktor. Razvijeni modeli mogu se primjenjivati za procjenu svojstava novodizajniranih spojeva samo iz njihove molekularne strukture.

Ključne riječi

Metoda potpornih vektora, kritična svojstva, algoritam optimizacije Dragonfly, kvantitativni odnos struktura-svojstvo

^aLaboratory of Biomaterials and Transport Phenomena (LBMP), University of Médéa, Médéa, Alžir

^bUniversity of Bouira, Bouira, Alžir

^cDepartment of Chemistry, University of Blida1, Blida, Alžir

^dLaboratory Quality Control, Physico-Chemical Department, Antibiotical Sidal of Médéa, Alžir

Izvorni znanstveni rad
Prispjelo 19. rujna 2020.
Prihvaćeno 8. prosinca 2020.