Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 18 (1) 35-48: Spring 2021
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Structural Relationship Study of Octanol-Water Partition Coefficient of the Compounds in kesum Essential Oil Using GA-MLR and GA-ANN Methods

Atefehsadat Navabi, Tahereh Momeni Isfahani*

Department of Chemistry, Arak Branch, Islamic Azad University, Arak, Iran

Received May 2021; Accepted June 2021

ABSTRACT

Essential Oils are highly concentrated substances the subtle, aromatic and volatile liquids. The use of essential oils is largely widespread in foods, deodorants, pharmaceuticals, drinks, cosmetics, medicine and embalming antiseptics especially with aromatherapy becoming increasingly popular. The lipophilicity of an organic compound can be described by a partition coefficient, logP, which plays a significant role in drug discovery and compound design. A data set of 40 compounds in the essential oil of kesum was randomly divided into 3 groups: training, test and validation sets consisting of 70%, 15% and 15% of data point, respectively. A large number of molecular descriptors were calculated with Dragon software. The Genetic Algorithm - Multiple Linear Regressions (GA-MLR) and genetic algorithm - artificial neural network (GA-ANN) were employed to design the Quantitative Structure-Property Relationship (QSPR) models. The predictive powers of the QSPR model was discussed using Coefficient of determination (R^2), Absolute Average Deviation (AAD) and the Mean Squared Error (MSE). The R^2 and MSE values of the MLR model were calculated to be 0.734 and 0.194 respectively. The R^2 and MSE values for the training set of the ANN model were calculated to be 0.9905 and 2×10^{-4} respectively. Comparison of the results revealed that the application the GA-ANN method gave better results than GA-MLR method.

Keywords: QSPR; multiple linear regressions; artificial neural network; genetic algorithm; essential oils; octanol- water partition coefficient

INTRODUCTION

Essential Oils are highly concentrated substances the subtle, aromatic and volatile liquids extracted from the flowers, leaves, stems, seeds, bark and roots of herbs, bushes, shrubs and trees through distillation. Natural essential oils are usually mixtures of terpenoids, aromatic and aliphatic compounds such as alcohols, aldehydes, ketones, carboxylic acids, esters, lactones and sulfides. The use of essential oils is largely widespread in foods, deodorants, pharmaceuticals, drinks, cosmetics, medicine and embalming antiseptics especially with aromatherapy

*Corresponding author: t.momeni@iau-arak.ac

becoming increasingly popular.

Essential oils and their constituents are commonly known for their antibacterial, antifungal and

antiparasitic activity, and there are also reports on the antimycobacterial properties[1].

Kesum (*Polygonum minus*) is an aromatic plant commonly used in Malay delicacies. This plant produces essential oil containing high levels of aliphatic aldehydes [2]. Kesum leaves is applied to hair to remove dandruff, used in aroma therapy [3] and in the perfume industry [4]. This plant has also been reported to possess several pharmacological properties like antimicrobial activity [5], cytotoxic activity [6], antioxidant activity [7] and anticancer activity [8,9].

Lipophilicity, as the ability of a molecule to mix with an oily phase rather than with water, is usually measured as partition coefficient, P, between the two phases and is often expressed as the logarithm of the partition coefficient between n-octanol and water ($logP_{ow}$). This coefficient is inversely related to the solubility of a compound in water. $LogP_{ow}$ is commonly used in Quantitative structure-property/ activity relationships (QSPRs/QSARs) studies and drug design [10-13] since this property is related to absorption [14], distribution [15], metabolism [16], excretion [17], and toxicity [18].

The QSAR models included the octanol-water partition coefficient as the molecular property and quantum mechanical descriptors such as the energies of the highest occupied molecular orbital and the lowest unoccupied molecular orbital, E_{HOMO} and E_{LUMO} have been applied to predict the antimycobacterial activity of Twenty-five constituents of essential oils [19]

The relationship between the molecular structures of the essential oil compounds and their antifungal activity have been done using the partial least squares (PLS) method ,$logP_{ow}$, E_{HOMO}, and the number of hydrogen-bond donor atoms in the molecules of the compounds studied (Donor) as molecular descriptors [20].

The QSAR studies have been widely used to understand the relationship between the chemical structure and biological activity of the molecules [19, 21].

The antibacterial activity of phenolic compounds in essential oils has been investigated by QSAR studies. These studies have been shown the importance of the contribution of the octanol-water partition coefficient (P_{ow}) in relation with the hydrophobic and amphiphilic character of the molecule [1].

Properties such as the n-octanol–water partition coefficient, Vapor pressures (P_V) and aqueous solubility ($S_{w,L}$) are important in predicting the environmental fate of organic compounds[22,23].

The objective of this study was to develop QSAR models for prediction the log P_{ow} of 40 compounds in kesum essential oil.

The QSPR model was constructed using the genetic algorithm (GA) variable selection, multiple linear regression and the Back-Propagation artificial neural network (BPANN) methods.

MATERIALS AND MATHEMATICAL METHODS

The chemical compounds in Essential Oils are compounds with a wide range of biological activities and they are the basis of several groups of drugs. A data set containing 40 compounds in kesum essential oil was used in this study. The chemical structure of molecules was drawn with the Gauss view program and optimized with the Gaussian 09W program based on the B3LYP functional and a 6-

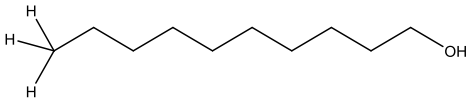
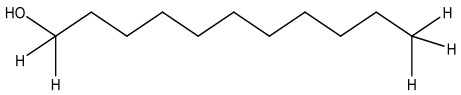
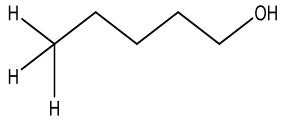
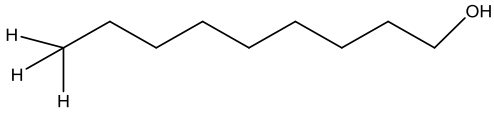
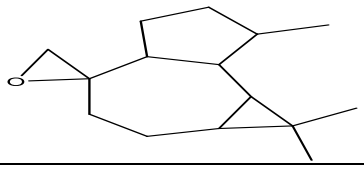
31G* basis set. The name, structure, formula and $logP_{ow}$ of the studied compounds in kesum, such as Nerolidol, Xanthorrhizol, Valencene, Farnesol and Drimenol are listed in Table 1. The $logP_{ow}$ values were taken from the literature [24].

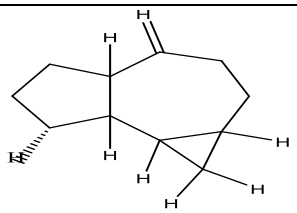
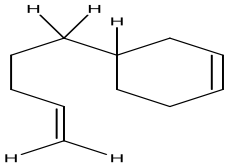
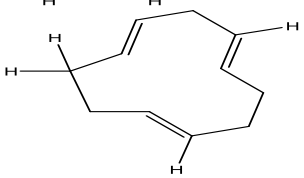
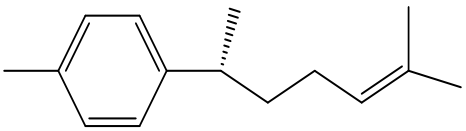
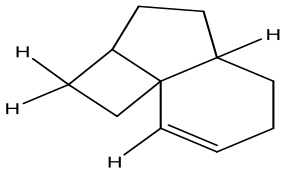
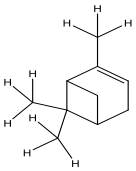
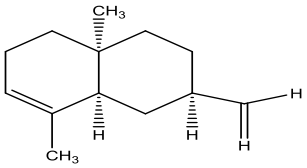
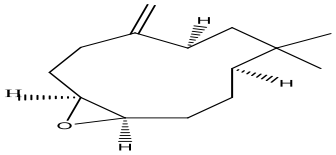
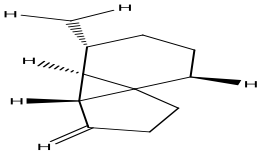
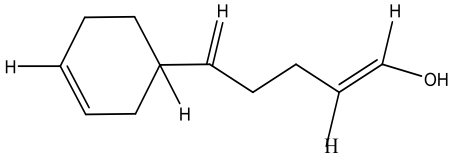
The molecular descriptor is the final result of logic and mathematical procedure which transforms chemical information encoded within a symbolic representation of a molecule into a useful number [25] and they are the independent chemical information used to predict the properties/activities of compounds in the research fields of QSPR/QSAR [26-28].

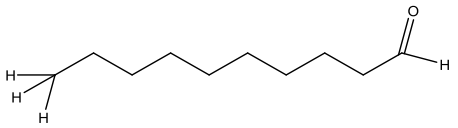
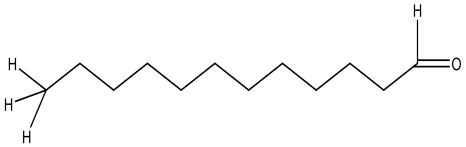
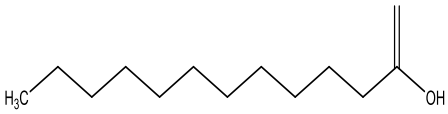
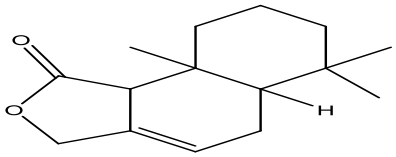
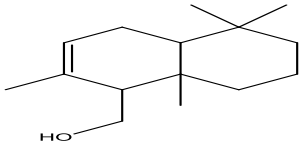
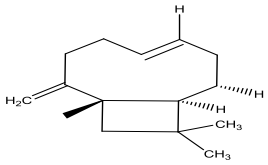
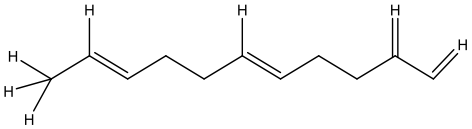
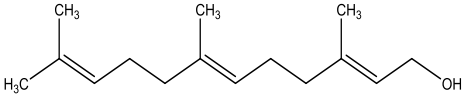
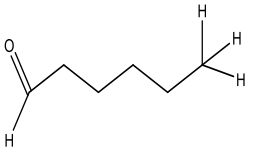
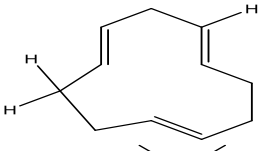
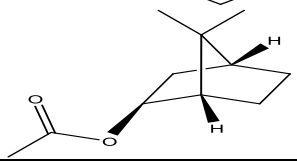
1489 molecular descriptors were calculated for selected compounds using the software DRAGON Version - 2006 package [29]. This software provides more than 4000 molecular descriptors that are divided into 20 logical blocks such as geometrical, getaway, WHIM, RDF, topological, functional group and constitutional descriptors [30, 31].

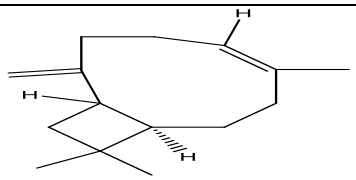
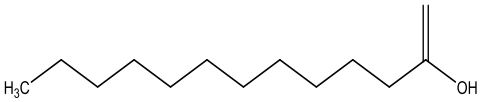
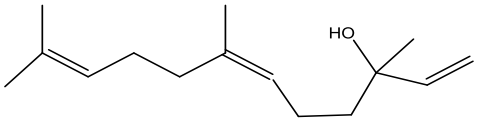
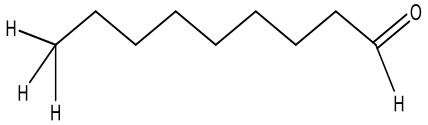
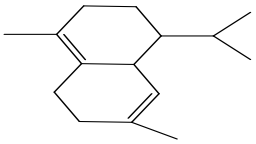
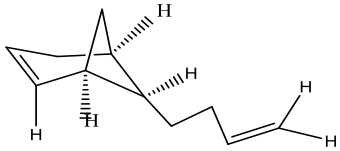
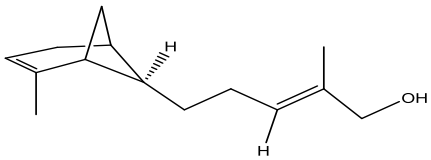
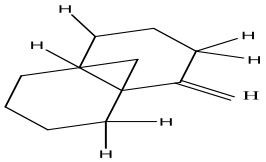
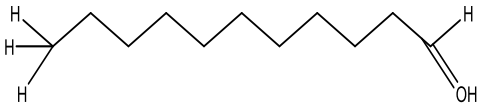
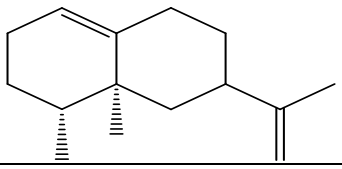
The genetic algorithm (GA) is written in MATLAB (version 2010a) environment has been used to reduce the number of descriptors derived from output Dragon software. Also backward stepwise regressions have been used to decrease the number of descriptors. The present back step program uses a backward variable-selection algorithm that starts with a set of n variables and, on the basis of a statistical usefulness criterion, selectively deletes one variable at a time to form progressively smaller subsets of predictors. After , remove the predictor with the highest p-value greater than 0.05. Then Refit the model and go to the Previous step. Stop when all p-values are less than 0.05. The software package SPSS 21.0 for Windows is used to implement multilinear regression [32].

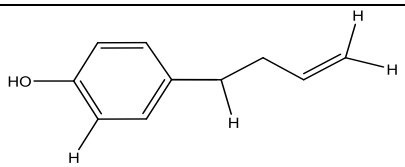
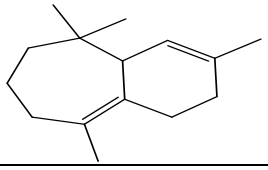
Table 1. The name, chemical structure of 40 compounds in kesum essential oils and their $logP_{ow}$ used in the present study

No.	Name	Formula	Log P_{OW}	Chemical structure
1	1-Decanol	$C_{10}H_{22}O$	3.12	
2	1-Dodecanol	$C_{12}H_{26}O$	3.9	
3	1-Hexanol	$C_6H_{14}O$	1.56	
4	1-Nonanol	$C_9H_{20}O$	2.73	
5	Alloaromadendrene oxid-(1)	$C_{15}H_{24}O$	3.48	

No.	Name	Formula	Log P_{OW}	Chemical structure
6	Alloaromadendrene	$C_{15}H_{24}$	4.27	
7	α-Bisabolol	$C_{15}H_{26}O$	3.65	
8	α-Caryophyllene	$C_{15}H_{24}$	4.89	
9	σ-Curcumene	$C_{15}H_{22}$	4.84	
10	*(-)*-σ-Panasinsene	$C_{15}H_{24}$	4.11	
11	α-Pinene	$C_{10}H_{16}$	3	
12	α-Selinene	$C_{15}H_{24}$	4.73	
13	β-Carvophyllene oxide	$C_{15}H_{24}O$	3.94	
14	β-Cubebene	$C_{15}H_{24}$	4.27	
15	cis-Lanceol	$C_{15}H_{24}O$	4.01	

No.	Name	Formula	Log P_{OW}	Chemical structure
16	Decanal	$C_{10}H_{20}O$	3.33	
17	Dodecanal	$C_{12}H_{24}O$	4.11	
18	Dodecanoic acid	$C_{12}H_{24}O_2$	3.99	
19	Drimenin	$C_{15}H_{22}O_2$	3.32	
20	Drimenol	$C_{15}H_{26}O$	3.78	
21	*(E)*-Caryophyllene	$C_{15}H_{24}$	4.73	
22	Farnesene	$C_{15}H_{24}$	5.2	
23	Farnesol	$C_{15}H_{26}O$	4.4	
24	Hexanal	$C_6H_{12}O$	1.77	
25	Humulene	$C_{15}H_{24}$	6.59	
26	Isobornyl acetate	$C_{12}H_{20}O_2$	2.76	

No.	Name	Formula	Log P_{OW}	Chemical structure
27	*iso*-Caryophyllene	$C_{15}H_{24}$	3.94	
28	*n*-Decanoic acid	$C_{10}H_{20}O_2$	3.21	
29	Nerolidol	$C_{15}H_{26}O$	4.4	
30	Nonanal	$C_9H_{18}O$	2.94	
31	*σ*-Cadinine	$C_{15}H_{24}$	4.74	
32	Tetradecanal	$C_{14}H_{28}O$	4.89	
33	*trans*-*α*-bergamotene	$C_{15}H_{24}$	4.73	
34	*trans*-*α*-*(Z)*-bergamotol	$C_{15}H_{24}O$	4.73	
35	*trans*-Longipinocarveol	$C_{15}H_{24}O$	3.53	
36	Undecanal	$C_{11}H_{22}O$	3.72	
37	Undecane	$C_{11}H_{24}$	4.54	
38	Valencene	$C_{15}H_{24}$	4.73	

No.	Name	Formula	Log P_{OW}	Chemical structure
39	Xanthorrhizol	$C_{15}H_{22}O$	4.55	
40	β-Himachalene	$C_{15}H_{24}$	4.08	

RESULTS AND DISCUSSION

Multiple Linear Regressions

The data set of 40 compounds was randomly separated into two groups, a training set of 30 compounds (75%) that was applied to create model and a test set of 10 compounds (25%) that was used to assess the performance of the made model.

Structural-activity model was generated using the backward multiple linear regression (BW-MLR) procedure of SPSS. The octanol-water partition coefficient ($logP_{ow}$) as the dependent variable and dragon molecular descriptors as the independent variable was used. Quality of the models was indicated by statistics parameters: correlation coefficient (R), squared regression coefficient (R^2), the Root Mean Squared Error (RMSE), Fisher ratio (F), Durbin- Watson (DW) and Significance (Sig) [33, 34].

The BW–MLR analysis led to the derivation of 6 models for the $logP_{ow}$, with 3-8 descriptors (Table 2). As can be seen, the three descriptors are useful to predict the $logP_{ow}$ which are: Mor09u (signal 09 / unweighted), ALOGP (Ghose-Crippen octanol-water partition coeff. (logP)) and EEig05x (Eigenvalue05 from edge adj,matrix weighted edge degrees). These descriptors are classified as 3D-MoRSE descriptors, molecular properties and Eigenvalues indices respectively.

With the selected descriptors, we have built the linear model using the training set data, (30 compounds) and obtained the following equation:

$$\log P_{OW} = 1.156 + 0.670\ \text{ALOGP} + 0.493\ \text{EEig05x} - 0.643\ \text{Mor09u} \quad (1)$$

N=30, R=0.857, R^2=0.734, R^2_{adj} =0.704, F=23.944, DW=1.808, Sig=0.000, MSE= 0.192

The MSE of this model was 0.192 and the R^2 value was 0.734. If squared regression coefficient is higher than 0.75 (R^2 >0.75), it indicates that there is a linear regression relationship between variables. But the MSE value indicated that the statistical results are not very satisfactory and there is no suitable linear relationship between molecular descriptors and octanol-water partition coefficient.

Table 2. Statistical parameters of the models calculated with the SPSS software for the log P_{OW}

models	Indepdndent Variables	R	R^2	R^2adj	RMSE	F	Sig
1	Mor25v,Mor09u,Mor31e,IDDE,ALOGP,X0Av,EEig05x,VEp1	0.893	0.798	0.721	0.438	10.350	0.000
2	Mor25v,Mor09u,Mor31e,IDDE,ALOGP,EEig05x,VEp1	0.892	0.796	0.731	0.438	12.257	0.000
3	Mor09u,Mor31e,IDDE,ALOGP, EEig05x,VEp1	0.887	0.787	0.731	0.438	14.140	0.000
4	Mor09u,Mor31e, ALOGP, EEig05x,VEp1	0.876	0.767	0.718	0.438	15.801	0.000
5	Mor09u, ALOGP, EEig05x,VEp1	0.863	0.746	0.705	0.438	18.309	0.000
6	Mor09u, ALOGP, EEig05x	0.857	0.734	0.704	0.438	23.944	0.000

Artificial Neural Network (ANN)

Because of the poor statistical results of the linear model, a non-linear model was also constructed in this study.

Artificial Neural Network (ANN) is an intelligence model and it imitates the working method of the human brain. A typical ANN system consists of a number of simple processing elements called neurons or nodes. These neurons are organized into different groups which are called layers. ANN contains three different layers: an input layer, one or more hidden layer, and an output layer of neurons [35].

In the ANN study, for the learning process, the data sets were randomly divided into 3 groups: training, test and validation sets consisting of 70%, 15% and 15% of the data point, respectively. The software would use the training data to build a basic model. The best algorithm based on minimum absolute error was selected when simulation trainings were completed [36, 37].

Among the ANN learning algorithms, the backpropagation (BP) method is one of the most generally used methods.

In this study, the BP algorithm strategy was used to develop and optimize the biases and the weights. The artificial neural network model is presented with Neural Network Toolbox techniques in MATLAB R2010b [38].

The number of input neurons was equal to that of the selected molecular descriptors. The GA-MLR selection procedure selected 3 descriptors for use as the input layers for ANN. The number of hidden neurons is an important parameter influencing the performances of the ANN model. In this work, we constructed BPANN model with 2-10 neurons in the hidden layer, individually and one node in the output layer. The input and output data were normalized between 0.1 and 0.9 using the following equation to avoid numerical overflows due to very large or very small weights.

$$y = 0.8 \times \left(\frac{x_i - x_{min}}{x_{max} - x_{min}} \right) + 0.1 \qquad (2)$$

The mean squared error (MSE) and squared regression coefficient (R^2) were calculated and recorded after every 10 cycles. The hidden layer with 6 neurons was produced the lowest MSE and the highest R^2.

(Fig. 1) shows the structure of a Back-Propagation Artificial Neural Network (BPANN).

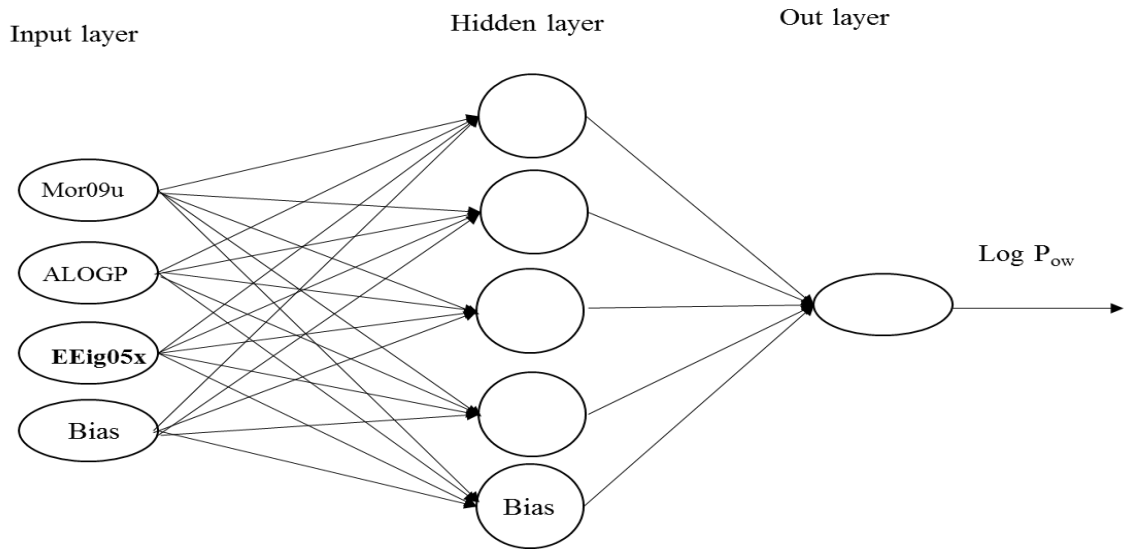


Fig. 1. Structure of a back-propagation artificial neural network.

To evaluate the ANN performance, the squared regression coefficients (R^2) Root Mean Square Error (RMSE) and Absolute Average Deviation (AAD) were used as criteria.

These parameters are defined as follows: The coefficient of determination (R^2) represents the fraction of the variance of Y "explained" by the correlation of Y with X. It gives the proportion of the variance (fluctuation) of one variable that is predictable from other variables [39].

$$R^2 = 1 - \sum_{i=1}^{n} \left\{ \frac{\left(y_{i,pred} - y_{i,exp}\right)^2}{\left(y_{avg,exp} - y_{i,exp}\right)^2} \right\} \quad (3)$$

The mean squared error (MSE) is defined as the average of the squares of the errors and the difference between the attribute which is to be estimated and the estimator [40], Root Mean Square Error (RMSE) is known to between observed and predicted estimated data is evaluated. Also, it is supposed that the indices with less estimated errors are more important [41].

$$MSE = \frac{1}{n}\sum_{i=1}^{i=N}\left(y_{i,pred} - y_{i,exp}\right)^2 \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(y_{i,pred} - y_{i,exp}\right)^2}{n}} \quad (5)$$

The absolute average deviation (AAD (%)) indicates the relative absolute deviation in percent from the calculation values.

$$ADD\% = \left\{ \frac{1}{n}\sum_{i=1}^{N}\left(\frac{\left(y_{i,pred} - y_{i,exp}\right)^2}{y_{i,pred}} \right) \right\} \times 100 \quad (6)$$

In the above equations, n is the number of experimental data; $y_{i,pred}$ and $y_{i,exp}$ are the predicted and experimental responses, respectively, and $y_{avg,\ exp}$ is the average of experimental values.

The R^2, MSE, RMSE and ADD values of total, training, testing and validation are listed in Table **3**.

Comparison of the values of MSE and other statistical parameters in Table **3** clearly indicates that the superiority of the GA-BPANN model over the GA-MLR model. The mean square error of 0.192 for the total set by the GA-MLR model should be compared with the value of 2×10^{-4} for the GA-BPANN model. Based on these results, there is the non-linear relationship between $logP_{ow}$ of the studied essential oils.

The experimental (observed) and predicted (calculated) values of the octanol-water partition coefficient of constituents of essential oils using BPANN and MLR models are listed in Table **4**.

Comparison of the residual values for $logP_{ow}$ of compounds in kesum essential oil versus the experimental values has been demonstrated in Fig.**3**. As can be seen the propagation of errors in both sides of zero are random shown in Fig. **3**.

Table 3. Performances of MLR and BPANN, QSAR Models

BPANN				
SET	R^2	MSE	RMSE	ADD
Total	0.9910	0.0002	0.0141	2.7358
Training	0.9905	0.0002	0.0151	2.7178
Test	0.9931	0.0043	0.0658	3.5902
Validation	0.9922	0.0002	0.0153	1.9652
MLR				
SET	R^2	MSE	RMSE	F
Total	0.734	0.192	0.438	23.94

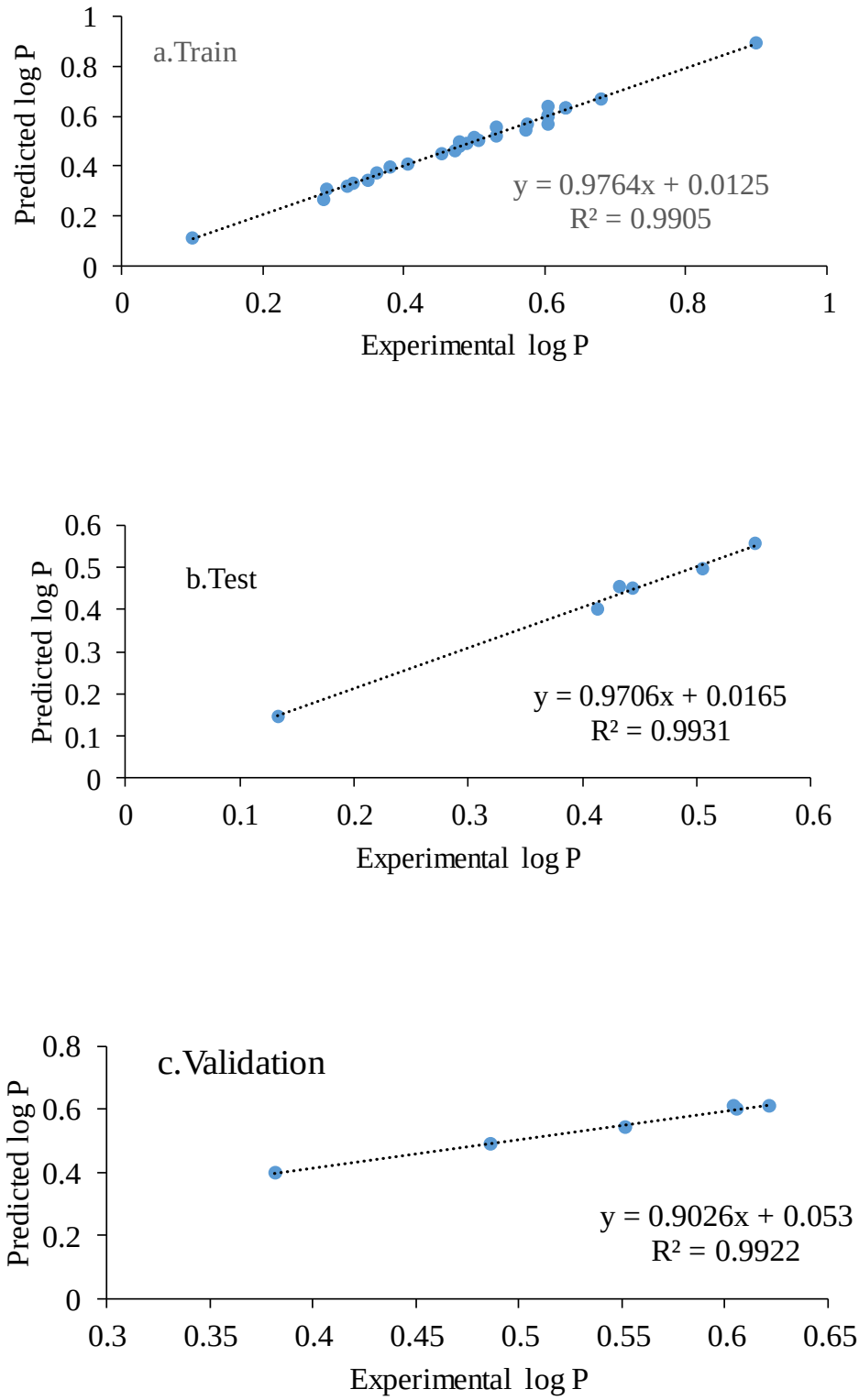


Fig. 2. Plot of the calculated logP against the experimental values for the training (a), test(b) and validation(c) sets.

Table 4. The experimental, predicted and residual values of the logP_{ow} by GA-BPANN and GA-MLR methods

	BPANN(normalized)			MLR		
No.	**Experimental log P_{OW}**	**Predicted log P_{OW}**	**Residual log P_{OW}**	**Experimental log P_{OW}**	**Predicted log P_{OW}**	**Residual log P_{OW}**
1	0.348	0.344	0.004	3.12	3.24639	-0.07895
2	0.472	0.462	0.010	3.9	3.74955	0.21975
3	0.100	0.114	-0.014	1.56	1.92783	0.00833
4	0.286	0.270	0.017	2.73	2.9486	-0.10925
5	0.405	0.406	-0.001	3.48	3.7262	-0.0434
6	0.531	0.557	-0.026	4.27	4.50305	0.09479
7	0.432	0.453	-0.021	3.65	4.32806	-0.09437
8	0.630	0.632	-0.003	4.89	4.51683	-0.05244
9	0.622	0.610	0.012	4.84	5.10338	0.06127
10	0.506	0.506	0.000	4.11	3.64512	-0.00592
11	0.329	0.329	0.000	3	2.93139	-0.00963
12	0.604	0.606	-0.002	4.73	4.51584	0.02276
13	0.479	0.481	-0.003	3.94	4.30821	0.01768
14	0.531	0.522	0.009	4.27	4.42634	0.06972
15	0.490	0.492	-0.002	4.01	4.09353	-0.03001
16	0.382	0.400	-0.018	3.33	3.37956	-0.02512
17	0.506	0.497	0.009	4.11	3.93631	-0.0684
18	0.486	0.492	-0.005	3.99	3.96594	0.00042
19	0.380	0.397	-0.017	3.32	3.54487	-0.07211
20	0.453	0.450	0.003	3.78	3.76043	-0.01019
21	0.604	0.611	-0.007	4.73	4.82554	0.2373
22	0.679	0.669	0.010	5.2	4.97812	0.10529
23	0.552	0.542	0.010	4.4	4.66123	-0.0708
24	0.133	0.146	-0.013	1.77	1.56476	0.04152
25	0.900	0.894	0.006	6.59	5.13554	0.02776
26	0.291	0.311	-0.020	2.76	2.55556	-0.01416
27	0.479	0.497	-0.018	3.94	4.86771	-0.3496
28	0.362	0.374	-0.011	3.21	3.34192	0.0673
29	0.552	0.555	-0.004	4.4	4.40545	-0.02304
30	0.319	0.318	0.001	2.94	2.92504	0.1459
31	0.606	0.599	0.007	4.74	4.32982	0.11172
32	0.630	0.632	-0.002	4.89	4.88552	0.04703
33	0.604	0.567	0.037	4.73	4.50621	-0.13859
34	0.604	0.595	0.009	4.73	3.97178	0.1654
35	0.413	0.402	0.012	3.53	3.47266	-0.00497
36	0.444	0.452	-0.009	3.72	3.77199	0.00185
37	0.574	0.546	0.028	4.54	4.17879	-0.16619
38	0.604	0.639	-0.035	4.73	4.65904	0.01669
39	0.576	0.570	0.006	4.55	5.17134	-0.11187
40	0.501	0.513	-0.012	4.08	4.44454	0.01655

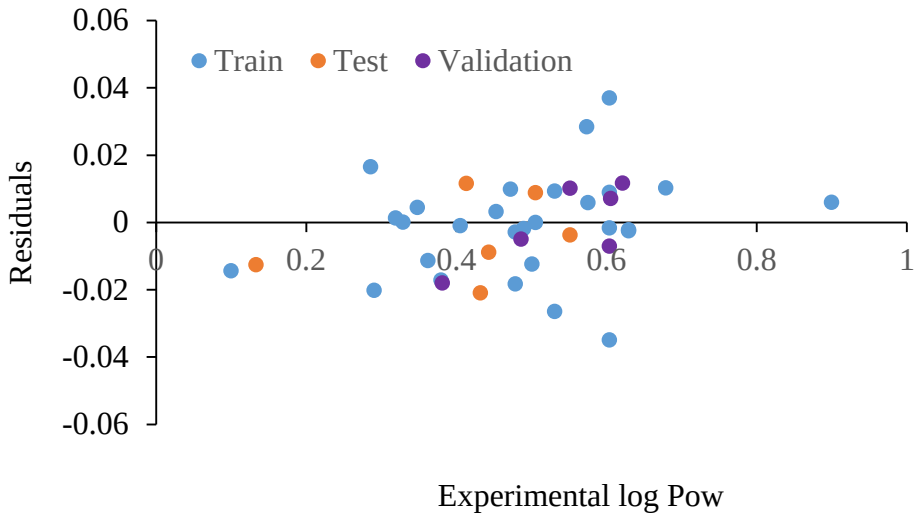


Fig.3. Plot of the residual values for $logP_{ow}$ of compounds *versus* the experimental values.

Interpretation of the best descriptors

The results and discussion lead us to conclude that combining of the three descriptors selected by GA, namely Mor09u, ALOGP and EEig05x can be used successfully for modeling and predicting the $logP_{ow}$ of compounds in kesum essential oil. These descriptors have been classified in 3D-MoRSE descriptor, molecular properties and Eigenvalues indices respectively. 3D-MoRSE descriptor derived from the knowledge of the 3D structure of the molecule and it plays a significant role in chemoinformatics and QSAR/QSPR purposes [42, 43]. Molecular properties indices have been improved such as Moriguchi logP, Ghose-Crippen logP, Lipinski rule-of-five, etc.

The GhoseCrippen octanol water coefficient (ALOGP) is a group contribution model for the octanolwater partition coefficient [44,45,46].

One of eigenvalue descriptor is the so-called eigenvalue-based topological molecular indices (EI). A descriptor from this set is defined using eigenvalues that come from one of the graph matrices (e.g. adjacency matrix). The EI introduced by Ernesto Estrada (therefore named as Estrada index) It has been successfully applied in modeling the folding in biomolecules. These indices can be classified into several groups by the nature of graph parameters used in their definitions [47-50].

CONCLUSIONS

In the present study, QSAR models have been developed to predict the $logP_{ow}$ of 40 compounds in kesum essential oil by genetic algorithm -multiple linear regression GA-MLR) and genetic algorithm - Back-Propagation Artificial Neural Network (GA-BPANN). Molecular descriptors were calculated with Dragon software and The Genetic Algorithm (GA) and backward Multiple Linear Regression (MLR) methods were used to select the suitable descriptors and to generate the correlation models that relate the chemical structural features to the biological activities.

The squared correlation coefficient (R^2), and mean square errors (MSE) have been designed to evaluate the quality and predictive ability of the linear and nonlinear models. Also other statistical

Parameters such as root mean squared error (RMSE), and absolute average deviation (AAD(%)) were used as a criterion. The R^2 and MSE values of the MLR and ANN models were calculated 0.734, 0.192and 0.9910, 2×10^{-4} respectively. The obtained results showed that the BPANN model with three selected descriptors (Mor09ul, ALOGP and EEig05x) could be used to predict $logP_{ow}$ of compounds in kesum essential oil.

REFERENCES

[1] S. Andrade-Ochoa, G. Virginia Nevárez-Moorillón, L. E. Sánchez-Torres, M. Villanueva-García, B. E. Sánchez-Ramírez, L. María Rodríguez-Valdez and B. E. Rivera-Chavira, BMC Compl Alternative Med. 15 (2015) 1.

[2] M. C. Gor, I. Ismail, W. A. W Mustapha, Z. Zainal, N. M. Noor and R. Othman, Acta. Physiol. Plant. 33(2011) 283.

[3] A. A. A. Almey, C. A. J. Khan, CAJ, I. S. Zahir, K. M. Suleiman, M. R. Aisyah and K. K. Rahim, Int. Food. Res. J. 17(2010) 1077.

[4] H. Bunawan, N. Talip and N. M. Noor, Aust. J. Crop. Sci. 5(2011) 123.

[5] A. M. Uyub, I. N. Nwachukwu, A. A. Azlan and S. S. Fariza, Res. Appl. 8(2010) 95.

[6] M. M. Mackeen, A. M. Ali, S. H. El-Sharkawy, M. Y. Manap, K. M. Salleh and N. H. Lajis, Pharm. Biol. 35(1997) 174.

[7] N. Huda-Faujan, A. Noriham, A. S. Norrakiah and A. S. Babji, J. Biotechnol. 8 (2009)484.

[8] A. M. Ali, M. M. Mackeen S. H. El-Sharkawy, J. Abdul Hamid, N. H. Ismail and F. Ahmad, Pertanika. J. Trop. Agric.Sci. 19(1996) 129.

[9] M. F. H´erent, V. D. Bie and B. Tilquin, J. Pharm. Biomed. Anal. 43(2007) 886.

[10] S. Ekins, J. Pharmacol. Toxicol. Methods. 53(2006) 38.

[11] J. A. Arnott, J. A. R. Kumar and S. L. Planey, J. Appl. Biopharm. Pharmacokinet. 1 (2013) 31.

[12] G. Korinth, T. Wellner, K. Schaller and H. Drexler, Toxicol. Lett. 215 (2012) 49.

[13] S. A .Wildman and G. M. Crippen, J. Chem. Inform. Comput. Sci. 39(1999) 868.

[14] A. Kokate, X. Li and B. Jasti, AAPS. Pharm.Sci.Tech. 9(2008) 501.

[15] S. Bharate, V. Kumar and R. A. Vishwakarma, Comb. Chem. High Throughput Screen. 19(2016) 461.

[16] H. A Zhong, V. Mashinson, T. A. Woolman and M. Zha, Curr. Top. Med. Chem. 13 (2013) 1290.

[17] S. Ito, H. Ando, A. Ose, Y. Kitamura, T. Ando, H. Kusuhara and Y. Sugiyama, J. Pharm. Sci.102 (2013) 3294.

[18] S. E. Czerwinski, J. P. Skvorak, D. M. Maxwell, D. E. Lenz and S.I. Baskin, J. Biochem. Mol. Toxic. 20 (2006) 241.

[19] S. Andrade-Ochoa, G. Virginia Nevárez-Moorillón, L. E. Sánchez-Torres, M. Villanueva-García, B. E. Sánchez-Ramírez, L. María Rodríguez-Valdez and B. E. Rivera-Chavira, BMC Complement.Altern. Med. 15 (2015)1.

[20] K. Voda, B. Boh Podgornik and M. Vrtačnik, J. Mol. *MODEL*.10(2004)76.

[21] R. Subramaniam and G. Rao, J. Comput. Methods. Mol. Des. 3 (2011) 69.

[22] H. Y. Xu, J. Y. Zhang, J. W. Zou and X. S. Chen, J. Mol. Graph. Model. 26 (2008) 1076.

[23] F. Wania and D. Mackay, Environ. Pollut. 100 (1999) 223.
[24] S. N. Baharum, H. Bunawan, M. A. A Ghani, W. A. W. Mustapha and N. M. Noor, Molecules. 15 (2010) 7006.
[25] R. Todeschini, M. Lasagni and E. Marengo, J. Chemometrics, 8 (1994) 263.
[26] H. Moriwaki, Y. S. Tian, N. Kawashita and T. Takagi, J. Cheminform. 10 (2018) 1.
[27] C. Yap, J. Comput. Chem. 32 (2010)1466.
[28] S. D. Bolboaca, L. Jantschi and M. V. Diudea, Curr. Comput. Aided. Drug. Des. 9 (2013) 195.
[29] Talete srl, Dragon (ver. 5.4), Milano, Italy. Web site: ww.talete.mi.it/products/software.htm
[30] R. Todeschini and V. Consonni, Handbook of molecular descriptors; Wiley-VCH: Weinheim, 2000.
[31] B. Bahadori and M. Atabati, Comb. Chem. High Throughput Screen. 20 (2017) 321.
[32] S. Ahmadi and E. Habibpour, Med. Chem. 17 (2017) 552.
[33] P. Kawczak, M. Belka, J. Slawinski and T. Baczek, Curr. Pharm. Anal. 14 (2018) 35.
[34] A. Rouhollahi, J. B. Ghasemi, E. Babaee and A. Ouammou, Curr. Anal. Chem. 6 (2010) 3.
[35] Y. Qudaih, T. Kerdphol and Y. Mitani, *SGCE*. 4 (2015) 247.
[36] S. Emamgholizadeh, M. Parsaeian and M. Baradaran, Europ. J. Agronomy. 68 (2015) 89.
[37] T. Zhou, S. Jhamb, X. Liang, K. Sundmacher and R. Gani, Chem. Eng. Sci.183(2018) 95.
[38] A. Thapliyal, R. Krishen Khar and A. Chandra, Curr. Nanosci. 14(2018) 239.
[39] R. G. D. Steel and J. H. Torrie, Principles and Procedures of Statistics with Special Reference to the Biological Sciences. McGraw Hill. 1960.
[40] E. L. Lehmann and G. Casella, Theory of Point Estimation (2nd ed.) New York, Springer. 1998.
[41] R. J. Hyndman and A. B. Koehler, Int. J. Forecast. 22 (2006) 679.
[42] O. Devinyak, D. Havrylyuk and R. Lesyk, J. Mol. Graph. Model. 54 (2014) 194.
[43] E. Dadfar, F. Shafiei and T. Momemi Isfahani, Curr Comput-Aid Drug. 1 (2019) 19.
[44] A. K Ghose, V. N. Viswanadhan and J. Wendoloski, J. Phys. Chem. 102(1998) 23772.
[45] V. Consonni, R. Todeschini, M. Pavan and P. Gramatica, J. Chem. Inform. Comput. Sci. 42 (2002) 693.
[46] F. Dialamehpour , F. Shafiei, J. Phys. Theor. Chem. 16(2020) 117
[47] X. Li, Y. Shi and I. Gutman I Graph energy. Springer, New York, 2012.
[48] M. V. Diudea, A. Pirvan–Moldovan, R. Pop and M. Medeleanu, MATCH. Commun. Math. Comput. Chem. 3 (2018)835.
[49] X. Ma, MATCH. Commun. Math. Comput. Chem. 81 (2019) 393.
[50] E. Estrada, Chem. Phys. Lett. 319 (2000) 713.

مجله شیمی فیزیک و شیمی نظری
دانشگاه آزاد اسلامی واحد علوم و تحقیقات
جلد ۱۸، شماره ۱، بهار ۱۴۰۰
ISSN ۱۷۳۵-۲۱۲۶

مطالعه ارتباط ساختار ضریب تقسیم اکتانول-آب در ترکیبات اسانس kesum با استفاده از روش‌های GA-MLR و GA-ANN

عاطفه سادات نوابی، طاهره مومنی اصفهانی*

گروه شیمی، واحد اراک، دانشگاه آزاد اسلامی اراک، اراک، ایران

چکیده

اسانس‌ها موادی مایع آروماتیک با غلظت بالا و معطر هستند. استفاده از اسانس‌ها در غذاها، دئودورانت ها، داروها، نوشیدنی-ها، لوازم آرایشی، داروها و ضد عفونی کننده‌های بسیار رایج هستند، به ویژه با رایحه درمانی که رو به افزایش است. چربی دوستی یک ترکیب آلی را می‌توان با ضریب توزیع، logP، که نقش مهمی در کشف دارو و طراحی آن دارد، توصیف کرد. ٤٠ ترکیب موجود در اسانس kesum به طور تصادفی به ۳ گروه تقسیم شدند: مجموعه آموزش، آزمون و اعتبارسنجی که به ترتیب شامل ۷۰٪ ، ۱۵٪ و ۱۵٪ از داده بود. تعداد زیادی توصیف کننده مولکولی با نرم‌افزار Dragon محاسبه شد. الگوریتم ژنتیک – رگرسیون خطی چندگانه (GA-MLR) و الگوریتم ژنتیک- شبکه عصبی مصنوعی (GA-ANN) برای طراحی مدل‌های کمی ساختار- ویژگی خواص (QSPR) استفاده شد. قدرت پیش‌بینی مدل QSPR با استفاده از ضریب تعیین (R^2)، انحراف متوسط مطلق (AAD) و خطای میانگین مربع (MSE) مورد بحث قرار گرفت. مقادیر R^2 و MSE مدل MLR به ترتیب۰/۷۳٤ و۰/۱۹٤ محاسبه شد. مقادیر R^2 و MSE برای مجموعه آموزشی مدل ANN به ترتیب ۰/۹۹۰۵ و ۲×۱۰^{-٤} محاسبه شد. مقایسه نتایج نشان داد که کاربرد GA-ANN نتایج بهتری نسبت به روش GA-MLR دارد.

کلید واژه‌ها: QSPR، رگرسیون خطی چند گانه، شبکه عصبی، الکوریتم ژنتیک، اسانس، ضریب توزیع آب- اکتانول

* مسئول مکاتبات: t.momeni@iau-arak.ac