



Available online at
www.heca-analitika.com/ijds

Infolitika Journal of Data Science

Vol. 1, No. 1, 2023



ANFIS-Based QSRR Modelling for Kovats Retention Index Prediction in Gas Chromatography

Rinaldi Idroes^{1,2,*}, Teuku Rizky Noviandy³, Aga Maulana³, Rivansyah Suhendra⁴, Novi Reandy Sasmita⁵, Muslem Muslem⁶, Ghazi Mauer Idroes⁷, Raudhatul Jannah³, Razief Perucha Fauzie Afidh³ and Irvanizam Irvanizam³

¹ Department of Chemistry, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; rinaldi.idroes@usk.ac.id (R.I.)

² Department of Pharmacy, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia;

³ Department of Informatics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; trizkynoviandy@gmail.com (T.R.N); agamaulana@usk.ac.id (A.M); raudhtl@gmail.com (R.J.); razief@usk.ac.id (R.P.F.A.); irvanizam.zamanhuri@usk.ac.id

⁴ Department of Information Technology, Faculty of Engineering, Universitas Teuku Umar, Aceh Barat 23681, Indonesia; rivansyahsuhendra@utu.ac.id (R.S.)

⁵ Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Syiah Kuala, Banda Aceh 23111, Indonesia; novireandys@usk.ac.id (N.R.S.)

⁶ Department of Chemistry, Faculty of Science and Technology, Universitas Islam Negeri Ar-Raniry, Banda Aceh 23111, Indonesia; moslem_coolam@yahoo.com (M.M)

⁷ Department of Occupational Health and Safety, Faculty of Health Sciences, Universitas Abulyatama, Aceh Besar 23372, Indonesia; idroesghazi_k3@abulyatama.ac.id (G.M.I.)

* Correspondence: rinaldi.idroes@usk.ac.id

Article History

Received 26 July 2023

Revised 26 August 2023

Accepted 4 September 2023

Available Online 7 September 2023

Keywords:

ANFIS

Essential oils

Molecular descriptors

Abstract

This study aims to evaluate the implementation and effectiveness of the Adaptive Neuro-Fuzzy Inference System (ANFIS) based Quantitative Structure Retention Relationship (QSRR) to predict the Kovats retention index of compounds in gas chromatography. The model was trained using 340 essential oil compounds and their molecular descriptors. The evaluation of the ANFIS models revealed promising results, achieving an R^2 of 0.974, an RMSE of 48.12, and a MAPE of 3.3% on the testing set. These findings highlight the ANFIS approach as remarkably accurate in its predictive capacity for determining the Kovats retention index in gas chromatography. This study provides valuable perspectives on the efficiency of retention index prediction through ANFIS-based QSRR methods and the potential practicality in compound analysis and chromatographic optimization.



Copyright: © 2023 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License. (<https://creativecommons.org/licenses/by-nc/4.0/>)

1. Introduction

The Kovats Retention Index is a dimensionless quantity used to express the relative retention time of a compound in a chromatographic column, normalized to the retention time of a reference compound [1]. This index is instrumental in identifying and comparing

compounds based on their elution order and retention behavior under specific chromatographic conditions. It serves as a crucial tool for researchers to understand how compounds interact within chromatography and make informed decisions about compound identification, purity, and composition [2].

The accurate prediction of Kovats RI holds immense significance in various scientific and industrial applications. It aids in compound identification, allowing researchers to match observed retention times with reference values, enabling reliable identification even in complex mixtures. Furthermore, Kovats RI prediction contributes to the optimization of chromatographic methods by assisting in selecting appropriate stationary phases, temperatures, and carrier gases, thereby enhancing separation efficiency and resolution.

Researchers have developed different methods for calculating the retention index in gas chromatography to achieve accurate results. These methods include the direct approach [3, 4], which uses the marker approach, and the indirect approach [5, 6], based on the homologous series approach and algorithm. Some of these techniques have shown good accuracy and practical applications, such as using iteration, spreadsheets, and statistical analysis. However, these methods still involve challenging laboratory procedures, particularly in obtaining precise standard series retention data, which is essential for accurate calculations [7].

In recent years, various modeling methods have been applied to predict the retention in gas chromatography. One of the popular and efficient methods among them is the Quantitative Structure-Retention Relationship (QSRR) [8, 9]. The key idea behind QSRR is to correlate the molecular descriptors of a compound with its retention index, enabling scientists to predict the retention behavior of new or unknown compounds without the need for extensive experimental measurements [10]. This approach is particularly valuable when dealing with a large number of compounds, as it saves time and resources compared to traditional trial-and-error experimental methods.

QSRR can be done using machine learning algorithms. In previous works, researchers have explored various approaches, including Multiple Linear Regression [11], Neural Network [12, 13], and Support Vector Regression [14, 15]. These methods have demonstrated good performance in establishing quantitative relationships between the chemical structure descriptors of compounds and their corresponding retention times or indices in gas chromatography.

Another potential method that can be used for QSRR is the Adaptive Neuro-Fuzzy Inference System (ANFIS), which was introduced by Jang in 1993 [16]. Many researchers have utilized this method to tackle various problems in the fields of science and engineering [17–20]. ANFIS is a hybrid computational model that combines

neural networks and fuzzy logic principles to create a powerful tool for modeling complex systems [21]. The advantages possessed by this method include its capability to adapt, fast learning capacity, and utilization of fuzzy logic, which allows ANFIS to handle uncertain or imprecise data, enhancing its robustness in dealing with real-world datasets [22].

This study addresses an important gap in gas chromatography, specifically in predicting Kovats retention indices for essential oils. While various predictive methods have been employed for this purpose, there is a need to explore alternative approaches that may offer improved accuracy and effectiveness. ANFIS, which has yet to be widely studied in this context, provides an opportunity to fill this gap and evaluate its potential for predicting Kovats retention indices. As a relatively less explored method in this context, ANFIS presents an opportunity to fill this gap and evaluate its potential as a robust tool for Kovats retention index prediction. We aim to contribute to the development of more precise retention index prediction methods within the field of gas chromatography, particularly for essential oil analysis.

2. Materials and Methods

2.1. Dataset

In this study, the dataset comprises a total of 340 compounds collected from the literature of Babushok [23]. These compounds serve the target variable, with their molecular properties represented by five distinct descriptors used as features. Molecular descriptors are mathematical representations of molecular properties generated by algorithms [24–26]. They numerically describe the physical and chemical information of a compound, enabling the prediction of the Kovats retention index. The molecular descriptors, namely ATSc1, VCH-7, SP-1, Kier1, and MLogP, were obtained from our previous works and have been identified as optimal choices based on selection through genetic algorithms. Table 1 provides a detailed overview of the molecular descriptors used in this study and their respective explanations.

To ensure a reliable and unbiased evaluation of the ANFIS model's predictive ability, we perform a dataset split into a training set and testing set, comprising 80% and 20% of the data, respectively [27, 28]. This partitioning is accomplished using the random splitting method, ensuring a representative distribution of data in both subsets. The visualization of the kernel density estimation for the training and testing sets is presented in Figure 1.

Table 1. Overview of the molecular descriptors in this study.

Descriptor Name	Definition
ATSc1	The Moreau-Broto autocorrelation descriptors with the incorporation of partial charges
VCH-7	Calculates the Kier & Hall Chi chain indices for orders 3, 4, 5, and 6.
SP-1	Computes the Kier & Hall Chi path indices with orders ranging from 0 to 7
Kier1	The descriptor used for computing Kier and Hall kappa molecular shape indices
MLogP	Moriguchi octanol-water partition coefficient calculation.

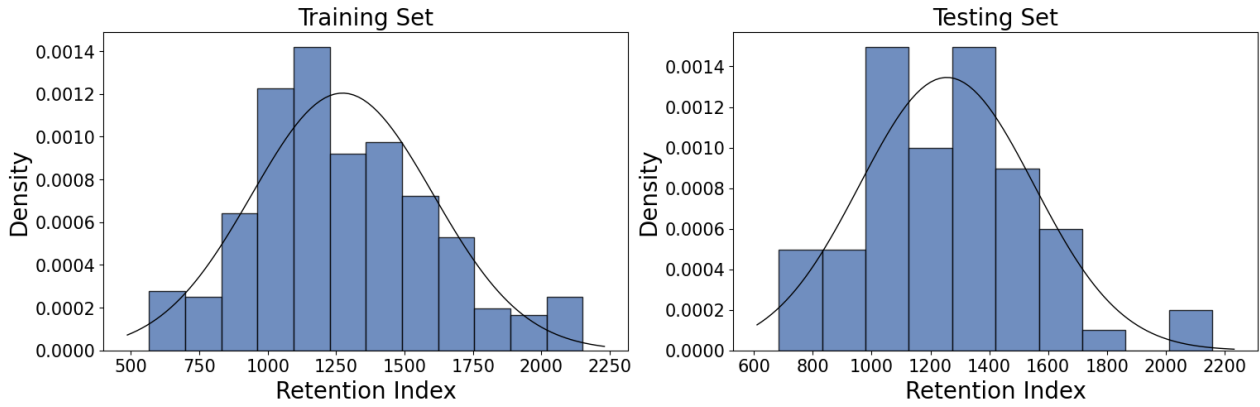


Figure 1. Kernel density estimation of the training and testing sets.

2.2 ANFIS

ANFIS is rooted in the theoretical foundation of fuzzy logic, designed to handle imprecise and uncertain information. Fuzzy logic employs linguistic variables and membership functions to represent and analyze vague data, thereby facilitating effective decision-making in uncertain environments [29]. ANFIS employs a hybrid learning algorithm that combines gradient descent and least squares estimation. This approach optimizes model parameters iteratively, refining fuzzy rules and membership functions to align with data patterns and enhance predictive accuracy. The ANFIS method consists of five layers. The explanation for each layer is as follows:

- In layer 1, there are nodes that provide output in the form of fuzzy membership grade from input with node functions:

$$O_i^1 = \mu_{A_i}(x), \quad (1)$$

where x is the input of node i , and μ_{A_i} is the linguistic label.

- In layer 2, there are nodes multiplying the incoming signal and sending the multiplication result as output:

$$O_i^2 = w_i = \mu_{A_i}(x) \times \mu_{A_i}(y), \quad i = 1,2 \quad (2)$$

The output of each node represents the firing strength of a rule (another T-norm operator that performs an AND operation can be used as a node function in this layer).

- In layer 3, there is an adaptive node/unfixed and labeled N. The output of the node in this layer is the result of calculating the ratio of the firing strength of the i -th rule to the total number of firing strengths in layer 2.

$$O_i^3 = \bar{w} = \frac{w_i}{w_1 + w_2}, \quad i = 1,2. \quad (3)$$

The output generated at this layer will be referred to as normalized fire strengths.

- In layer 4, the summation process is carried out at this layer:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i), \quad (4)$$

where \bar{w}_i is the output of layer 3, and $\{p_i, q_i, r_i\}$ is the set of parameters. The parameters in this layer will be called consequent parameters.

- In layer 5, there is one node labeled, which will calculate the overall result as the sum of all incoming signals.

$$O_i^5 = \sum_i \bar{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \quad (5)$$

We utilize a type-3 ANFIS model that follows the type-3 fuzzy reasoning method. The ANFIS model has two sets of parameters, namely premise and consequent parameters, which we update during the learning process. For parameter updates, we employ a hybrid

algorithm that combines least square estimation and gradient descent. Specifically, we update the consequent parameters during the forward step using least square estimation, while the backward step involves updating the premise parameters using gradient descent.

To construct the ANFIS model, we opt for the Gaussian membership function, employing two membership functions. To obtain the values for the Gaussian membership function parameters (i.e., mean and standard deviation), we cluster the dataset using the Fuzzy C-Means clustering method with two clusters. Thus, the mean and standard deviation of each of these clusters are utilized as the parameters for the Gaussian membership function [30].

2.3. Performance Evaluation

The evaluation of the model's predictive capabilities is conducted using various statistical metrics, namely the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute percentage error (MAPE). These metrics were employed together to gain comprehensive insights into the model's accuracy, goodness of fit, and ability to make reliable predictions for practical applications.

The equations for R^2 , RMSE, and MAPE are presented in equations 6, 7, and 8, respectively:

$$R^2 = \frac{\sum_{i=1}^n (y_i^{fit} - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \tag{7}$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - y_i^{fit}}{y_i} \right| \tag{8}$$

where y_i^{fit} is the fitted value, \bar{y} is the average of the observed values, y_i is the observed value, \hat{y}_i is the predicted value, and n is the amount of data. The acceptable range values for the evaluation metrics R^2 , RMSE, and MAPE are commonly used as guidelines to assess the performance of predictive models across various studies. R^2 values greater than 0.6 indicate a reasonably good fit of the model to the data, suggesting that a substantial proportion of the variance is captured. An RMSE less than 10% of the target value range suggests that the model's average prediction error is within an acceptable range. Likewise, a MAPE of less than 10% indicates that the model's predictions demonstrate a high level of accuracy relative to the observed values. [31, 32].

3. Results and Discussion

The ANFIS model was utilized in this study to predict the Kovats retention index of essential oils. Four iterations of the model were trained, each with a different learning rate (0.1, 0.01, 0.001, and 0.0001). We employed different learning rates in the training process to investigate the impact of learning rates on the performance of the ANFIS model in predicting the Kovats retention index of essential oils. The learning rate is a hyperparameter that controls the step size of the gradient descent optimization algorithm during training. A higher learning rate allows the model to update its parameters more quickly, but it may risk overshooting the optimal solution and result in unstable convergence. Conversely, a lower learning rate leads to more cautious parameter updates, potentially leading to slower convergence but with better stability. All models underwent 50 epochs of training. Then, we evaluated their performance on both the training set and the testing set. Evaluating the training set allows us to assess how well the models have learned the training data, providing insight into their ability to fit the training data accurately. On the other hand, evaluating the testing set serves as an important step to assess the generalization capability of the models, indicating how well they can make predictions on unseen data and highlighting any overfitting issues that might have arisen during the training process. The results of this evaluation are presented in Table 3.

The model with a learning rate of 0.001 stands out by achieving the highest R^2 score (0.986), along with the lowest RMSE (39.37) and MAPE (2.5%) on the testing set. These results underline its superior predictive capability in comparison to other learning rates. Several key factors contribute to this outcome. Primarily, the adoption of smaller steps during parameter updates promotes a gradual convergence toward the optimal solution, fostering stability throughout the training process. This measured approach minimizes the risk of overshooting optimal values, thus maintaining a consistent and accurate convergence trajectory.

Moreover, the lower learning rate enhances the model's efficiency in traversing the intricate loss landscape, resulting in improved convergence accuracy. Another advantage lies in the ability of smaller learning rates to prevent the model from becoming trapped in local minima, thereby encouraging exploration within the optimization space. This cautious parameter update strategy effectively mitigates divergence, a common concern associated with higher learning rates. Furthermore, the fine-tuning of model weights is refined due to the precise adjustments offered by the reduced

Table 3. Performance evaluation of the ANFIS model.

Learning Rate	Training Set			Testing Set		
	R ²	RMSE	MAPE	R ²	RMSE	MAPE
0.1	0.983	42.57	2.7%	0.958	60.43	3.8%
0.01	0.981	45.49	2.9%	0.962	57.73	3.6%
0.001	0.986	39.79	2.6%	0.974	48.12	3.3%
0.001	0.986	39.37	2.5%	0.963	56.82	3.8%

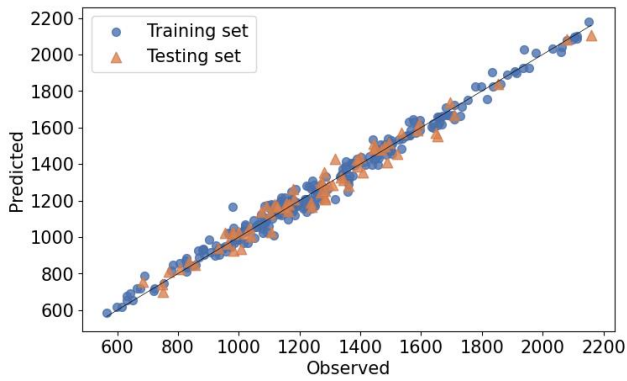


Figure 2. Observed vs. predicted of the ANFIS model.

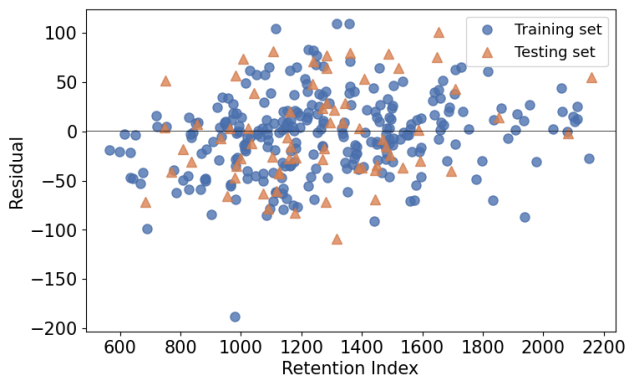


Figure 3. Residual plot of the ANFIS model.

learning rate. The associated smoother learning curves contribute to comprehensive training progress monitoring, promptly detecting any irregularities.

The results also show that the model trained with a learning rate of 0.1 resulted in the worst performance on both the training and testing sets, likely because the learning rate is too high for our dataset, causing the model to diverge and making it difficult for the optimization algorithm to find the true minimum of the loss function. However, all ANFIS models demonstrated satisfactory predictive capability, and even the weaker performers yielded results within an acceptable range for practical applications.

Importantly, the lower learning rate adeptly manages data and feature sensitivity, preventing overfitting and accommodating varying feature magnitudes. Lastly, the implicit regularization effect exhibited by lower learning rates steers the model towards capturing meaningful

data patterns while resisting noise influence. However, the specific learning rate selection should be balanced based on the trade-off between convergence speed and stability to ensure the model's optimal performance in predicting the Kovats retention index.

The results also show that the model trained with a learning rate of 0.1 resulted in the worst performance on both the training and testing sets, likely because the learning rate is too high for our dataset, causing the model to diverge and making it difficult for the optimization algorithm to find the true minimum of the loss function. However, all ANFIS models demonstrated satisfactory predictive capability, and even the weaker performers yielded results within an acceptable range for practical applications.

To better understand the results achieved by the model when using a learning rate of 0.001, we are taking a closer look at its abilities through visualizations. Figure 2 shows the observed vs. predicted plot of the ANFIS model. It becomes clear that the data points are noticeably arranged close to a reference line. This line symbolizes an ideal scenario in which the observed and predicted values perfectly match. The proximity of the scattered data points to this reference line reflects how accurately the predictive model performs. When the data points are nearer to the line, it signifies a stronger agreement between the predicted outcomes and the actual observations. When data points are tightly clustered around the reference line, it indicates a positive outcome, implying that the predictive model has effectively captured the inherent patterns within the data. This alignment suggests that the model's predictions closely resemble the observed values, showcasing a dependable and robust predictive performance.

Subsequently, Figure 3 displays the residual plot generated by our ANFIS model. This plot provides insights into the discrepancies or differences between the predicted values and the actual observed values within the model. It can be seen that the ANFIS regression model exhibited varying levels of accuracy across the dataset. Some data points show relatively small residuals, suggesting a close match between predicted and actual values. Overall, a general trend is observed wherein the residuals tend to center around zero, which shows that

the ANFIS model was adept at capturing the overall patterns present in the dataset.

4. Conclusions

In this study, we employed an ANFIS model to predict the Kovats retention index of 340 essential oil compounds. The ANFIS-based models demonstrated a strong predictive capacity, with the best result achieving an R^2 of 0.974, an RMSE of 48.12, and an MAPE of 3.3% on the testing set. These results underscore the ANFIS's potential for accurate Kovats retention index prediction, making it a valuable tool for compound analysis and chromatographic process enhancement. The demonstrated accuracy and reliability of ANFIS highlight its ability to expedite compound identification procedures, driving progress in scientific research and industrial applications. Future studies could consider expanding the dataset to cover a broader compound range and exploring the applicability of ANFIS for predicting additional chromatographic parameters.

Author Contributions: Conceptualization, R.I., T.R.N. A.M. and I.I.; methodology, R.I., T.R.N., R.S., N.R.S. and R.J.; software, A.M., R.J. and R.P.F.A.; validation, R.I., M.M., G.M.I. and I.I.; formal analysis, R.I., R.S., G.M.I. and R.P.F.A.; investigation, R.I., T.R.N., M.M. and R.J.; resources, R.I. and G.M.I.; data curation, R.I., R.P.F.A. and I.I.; writing—original draft preparation, R.I., T.R.N., A.M. and R.J.; writing—review and editing, R.I., R.P.F.A. and I.I.; visualization, N.R.S. and G.M.I.; supervision, R.I.; project administration, R.I.; funding acquisition, R.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Universitas Syiah Kuala, Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi through the “Penelitian Profesor” scheme (grant number: 166/UN11/SPK/PNBP/2021).

Ethical Clearance: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data is available upon request to corresponding author.

Acknowledgments: The authors extend their sincere gratitude to their respective universities for their support and encouragement throughout the course of this study.

Conflicts of Interest: The author declares that there is no conflict of interests regarding the publication of this manuscript.

References

- Zenkevich, I. G. (2010). Kovats' Retention Index System, *Encyclopedia of Chromatography*, Vol. 2, 1304–1310.
- Anjum, A., Liigand, J., Milford, R., Gautam, V., and Wishart, D. S. (2023). Accurate Prediction of Isothermal Gas Chromatographic Kovats Retention Indices, *Journal of Chromatography A*, 464176.
- Korány, K., and Amtmann, M. (2005). A practical, theory-supported approach of linear temperature programmed gas chromatographic retention indices used in the recognition experiments of Hungarian food specialities, called “Hungarics”, *Journal of Food Composition and Analysis*, Vol. 18, No. 5, 345–357.
- Wiedmer, S. K., Lokajova, J., and Riekkola, M. (2010). Marker compounds for the determination of retention factors in EKC, *Journal of Separation Science*, Vol. 33, No. 3, 394–409.
- Idroes, R., Muslem, Mahmudi, Saiful, Idroes, G. M., Suhendra, R., and Irvanizam. (2020). The effect of column and temperature variation on the determination of the dead time in gas chromatographic systems using indirect methods, *Heliyon*, Vol. 6, No. 2, e03302–e03302. doi:10.1016/j.heliyon.2020.e03302.
- Aalizadeh, R., Nikolopoulou, V., and Thomaidis, N. S. (2022). Development of Liquid Chromatographic Retention Index Based on Cocamide Diethanolamine Homologous Series (C (n)-DEA), *Analytical Chemistry*, Vol. 94, No. 46, 15987–15996.
- Idroes, R., Mahmudi, M., Saiful, S., Muslem, M., Idroes, G. M., Rusyana, A., Suhendra, R., and Irvanizam, I. (2021). SPREADSHEET AND ITERATION DEAD TIMES OF 2- ALKANONES HOMOLOGUES SERIES ELUTED BY ACETONITRILE/WATER IN REVERSED PHASE HIGHPERFORMANCE LIQUID CHROMATOGRAPHY, *Rasayan Journal of Chemistry*, Vol. 14, No. 02, 1041–1048. doi:10.31788/RJC.2021.1426024.
- Kumar, A., Kumar, P., and Singh, D. (2022). QSRR modelling for the investigation of gas chromatography retention indices of flavour and fragrance compounds on Carbowax 20 M glass capillary column with the index of ideality of correlation and the consensus modelling, *Chemometrics and Intelligent Laboratory Systems*, Vol. 224, 104552. doi:10.1016/j.chemolab.2022.104552.
- Aćimović, M., Pezo, L., Tešević, V., Čabarkapa, I., and Todosijević, M. (2020). QSRR Model for predicting retention indices of Satureja kitaibelii Wierzb. ex Heuff. essential oil composition, *Industrial Crops and Products*, Vol. 154, 112752. doi:10.1016/j.indcrop.2020.112752.
- Liapikos, T., Zisi, C., Kodra, D., Kademoglou, K., Diamantidou, D., Begou, O., Pappa-Louisi, A., and Theodoridis, G. (2022). Quantitative structure retention relationship (QSRR) modelling for Analytes' retention prediction in LC-HRMS by applying different Machine Learning algorithms and evaluating their performance, *Journal of Chromatography B*, Vol. 1191, 123132. doi:10.1016/j.jchromb.2022.123132.
- Idroes, R., Noviandy, T. R., Maulana, A., Suhendra, R., Sasmita, N. R., Muslem, M., Idroes, G. M., Kemala, P., and Irvanizam, I. (2021). Application of Genetic Algorithm-Multiple Linear Regression and Artificial Neural Network Determinations for Prediction of Kovats Retention Index, *International Review on Modelling and Simulations (IREMOS)*, Vol. 14, No. 2, 137. doi:10.15866/iremos.v14i2.20460.
- Qu, C., Schneider, B. I., Kearsley, A. J., Keyrouz, W., and Allison, T. C. (2021). Predicting Kovats Retention Indices Using Graph Neural Networks, *Journal of Chromatography A*, Vol. 1646, 462100. doi:10.1016/j.chroma.2021.462100.
- Maulana, A., Noviandy, T. R., Idroes, R., Sasmita, N. R., Suhendra, R., and Irvanizam, I. (2020). Prediction of Kovats Retention Indices for Fragrance and Flavor using Artificial Neural Network, *2020 International Conference on Electrical Engineering and Informatics (ICELTICs)*, IEEE, 1–5. doi:10.1109/ICELTICs50595.2020.9315391.
- Zhang, J., Zheng, C. H., Xia, Y., Wang, B., and Chen, P. (2017). Optimization enhanced genetic algorithm-support vector regression for the prediction of compound retention indices in gas chromatography, *Neurocomputing*, Vol. 240, 183–190. doi:10.1016/j.neucom.2016.11.070.
- Noviandy, T. R., Maulana, A., Sasmita, N. R., Suhendra, R., Irvanizam, I., Muslem, M., Idroes, G. M., Yusuf, M., Sofyan, H., Abidin, T. F., and Idroes, R. (2022). The Prediction of Kovats Retention Indices of Essential Oils at Gas Chromatography Using Genetic Algorithm-Multiple Linear Regression and Support Vector Regression, *Journal of Engineering Science and Technology*, Vol. 17, No. 1, 306–326.

16. Jang, J.-S. R. (1993). ANFIS: adaptive-network-based fuzzy inference system, *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 23, No. 3, 665–685. doi:10.1109/21.256541.
17. Hussain, M., Raju, Y. K., and V Kamakshi, P. (2021). Application of Multiple Linear Regression Models and Adaptive Neuro-Fuzzy Inference System Models to estimate the Compressive Strength of Concrete, *IOP Conference Series: Materials Science and Engineering*, Vol. 1126, No. 1, 012062. doi:10.1088/1757-899X/1126/1/012062.
18. Senthilraja, S., Gangadevi, R., Köten, H., and Marimuthu, R. (2020). Performance assessment of a solar powered hydrogen production system and its ANFIS model, *Heliyon*, Vol. 6, No. 10, e05271. doi:10.1016/j.heliyon.2020.e05271.
19. Yilmaz, I., and Kaynar, O. (2011). Multiple regression, ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils, *Expert Systems with Applications*, Vol. 38, No. 5, 5958–5966. doi:10.1016/j.eswa.2010.11.027.
20. Elaziz, M. A., Moemen, Y. S., Hassaniien, A. E., and Xiong, S. (2018). Quantitative Structure-Activity Relationship Model for HCVNS5B inhibitors based on an Antlion Optimizer-Adaptive Neuro-Fuzzy Inference System, *Scientific Reports*, Vol. 8, No. 1, 1506. doi:10.1038/s41598-017-19122-y.
21. Al-Hmouz, A., Jun Shen, Al-Hmouz, R., and Jun Yan. (2012). Modeling and Simulation of an Adaptive Neuro-Fuzzy Inference System (ANFIS) for Mobile Learning, *IEEE Transactions on Learning Technologies*, Vol. 5, No. 3, 226–237. doi:10.1109/TLT.2011.36.
22. Şahin, M., and Erol, R. (2017). A Comparative Study of Neural Networks and ANFIS for Forecasting Attendance Rate of Soccer Games, *Mathematical and Computational Applications*, Vol. 22, No. 4, 43. doi:10.3390/mca22040043.
23. Babushok, V. I. (2015). Chromatographic retention indices in identification of chemical compounds, *TrAC - Trends in Analytical Chemistry*, Vol. 69, 98–104. doi:10.1016/j.trac.2015.04.001.
24. Noviandy, T. R., Maulana, A., Emran, T. B., Idroes, G. M., and Idroes, R. (2023). QSAR Classification of Beta-Secretase 1 Inhibitor Activity in Alzheimer's Disease Using Ensemble Machine Learning Algorithms, *Heca Journal of Applied Sciences*, Vol. 1, No. 1, 1–7. doi:10.60084/hjas.v1i1.12.
25. Todeschini, R., and Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*. Vol. 1. Alphabetical Listing; Vol. 2. Appendices, References.
26. Maulydia, N. B., Khairan, K., and Noviandy, T. R. (2023). Prediction of Pharmacokinetic Parameters from Ethanolic Extract Mane Leaves (*Vitex pinnata* L.) in Geothermal Manifestation of Seulawah Agam Ie-Seu'um, Aceh, *Malacca Pharmaceutics*, Vol. 1, No. 1, 16–21. doi:10.60084/mp.v1i1.33.
27. Noviandy, T. R., Maulana, A., Idroes, G. M., Maulydia, N. B., Patwekar, M., Suhendra, R., and Idroes, R. (2023). Integrating Genetic Algorithm and LightGBM for QSAR Modeling of Acetylcholinesterase Inhibitors in Alzheimer's Disease Drug Discovery, *Malacca Pharmaceutics*, Vol. 1, No. 2, 48–54. doi:10.60084/mp.v1i2.60.
28. Joseph, V. R. (2022). Optimal ratio for data splitting, *Statistical Analysis and Data Mining: The ASA Data Science Journal*, Vol. 15, No. 4, 531–538. doi:10.1002/sam.11583.
29. Agustia, M., Noviandy, T. R., Maulana, A., Suhendra, R., Muslem, M., Sasmita, N. R., Idroes, G. M., Rahimah, S., Afidh, R. P. F., Subianto, M., Irvanizam, I., and Idroes, R. (2022). Application of Fuzzy Support Vector Regression to Predict the Kovats Retention Indices of Flavors and Fragrances, *2022 International Conference on Electrical Engineering and Informatics (ICELTICs)*, IEEE, 13–18. doi:10.1109/ICELTICs56128.2022.9932124.
30. Palconit, M. G. B., Conception II, R. S., Alejandrino, J. D., Nunez, W. A., Bandala, A. A., and Dadios, E. P. (2021). Comparative ANFIS Models for Stochastic On-road Vehicle CO₂ Emission using Grid Partitioning, Subtractive, and Fuzzy C-means Clustering, *2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC)*, IEEE, 1–6. doi:10.1109/R10-HTC53172.2021.9641644.
31. Alexander, D. L. J., Tropsha, A., and Winkler, D. A. (2015). Beware of R²: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models, *Journal of Chemical Information and Modeling*, Vol. 55, No. 7, 1316–1322. doi:10.1021/acs.jcim.5b00206.
32. Moreno, J. J. M., Palmer Pol, A., Sesé Abad, A., and Cajal Blasco, B. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy., *Psicothema*, Vol. 25, No. 4, 500–6. doi:10.7334/psicothema2013.23.