Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)
www.abjournals.org

CHEMOMETRICS FEASIBILITY FOR MODELLING CHROMATOGRAPHIC BEHAVIOR OF DIAZEPAM USING LINEAR AND NON-LINEAR TECHNIQUES: A DATA MINING BASED APPROACH

Mohamed Alhosen Degm[1], Mohamed Miftah Salem Ahmed[2], Zenib M. Zaydi[3], Mohamed Ab. Khalifa Ibrahim[4], Umar M. Ghali[5], Shamsu Shuaibu Bala[6], Aliyu Yakubu[7], and Bashir Ismail Ahmad[8*]

[1,4,5]Faculty of Science, Department of Chemisty, Çankırı Karatekin University, Turkey.

[2,3]High Institute of Science and Technology – Alqarahbuli, Libya.

[6]Rabiu Musa Kwankwaso College of Advanced and Remedial Studies, Tudun wada Kano, Nigeria.

[7]Department of biochemistry, Kaduna State university. Kaduna state. Nigeria.

[8]Department of Chemistry, Near East University, TRNC, Mersin 10, 99138 Nicosia, Turkey.

[*]Corresponding Author's Email: ahmadismailbashir@gmail.com

Cite this article:

Mohamed A. D., Mohamed M. A., Zenib M. Z., Mohamed A. I., Umar M. G., Shamsu S. B., Aliyu Y., Bashir I. A. (2024), Chemometrics Feasibility for Modelling Chromatographic Behavior of Diazepam using Linear and Non-Linear Techniques: A Data Mining Based Approach. Advanced Journal of Science, Technology and Engineering 4(1), 119-134. DOI: 10.52589/AJSTE-XTWJGQQL

Manuscript History

Received: 26 Feb 2024

Accepted: 6 May 2024

Published: 13 Jun 2024

ABSTRACT: *Modern data mining employs the use of statistics with other tools, ideas and approaches from computer science, database technology, machine learning, deep learning as well as other classical analytical tools. In this research, three distinct algorithms have been employed for modelling the performance properties of Diazepam (DIA) utilizing the method of high-performance liquid chromatography (HPLC). The algorithms developed include support vector machine (SVM), adaptive-neuro fuzzy inference systems (ANFIS), and linear regression (LR). Temperature and mobile phase inform of methanol (MeOH) were used as predictors, while the time recorded for the retention was deemed to be the dependent variable. The performance accuracy of the models was assessed using two statistical metrics, including determination co-efficient (R^2) and root mean square error (RMSE). The obtained results were shown both qualitatively and graphically using different charts. The comparative performance accuracy of the models demonstrates that the non-linear models (ANFIS and SVM) displays a higher performance efficiency than LR and has the ability of enhancing its performance ability by 51.2% and 76.1% both throughout the training and testing phase.*

KEYWORDS: Diazepam; HPLC; Chemometrics; Artificial intelligence; Data Mining.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)

www.abjournals.org

INTRODUCTION

The concept of data mining involves the discovery of various interesting, valuable and unexpected structures in either large or small datasets. Therefore, it composed of two different aspects (Abba et al., 2021; Abba & Usman, 2020; Sohrabi et al., 2019; A. G. Usman, Ahmad, et al., 2021). The first one involves the global structures, which aimed at modelling the distribution inform of regression, shapes, features present in the shape as well as classification (Rong et al., 2018). The second one involves the small scale or sometimes considered as the local structures, which mainly focused on modelling and detecting some anomalies as well as decide if there is high performance for fitness to decide whether there is chance for their occurrence (Bala et al., 2023; Rezaei-Darzi et al., 2014; A. G. Usman, IŞIK, et al., 2022). Data mining has been in existence for long, which is recently known as derogatory way, fishing through data and trawling (Abba et al., 2022; Abba, Benaafi, Usman, & Aljundi, 2023; Abba, Benaafi, Usman, Ozsahin, et al., 2023; Jibril, Zayyan, et al., 2023; Madaki et al., 2022; Pham et al., 2019; Rong et al., 2018; A. G. Usman, Ghali, et al., 2022; D. Uzun Ozsahin et al., 2022; D. U. Uzun Ozsahin et al., 2023; Yassin et al., 2022). In the concept of data mining in pharmaceutical sector, most of the analysis involves the second type of data mining; even though, the detection of various signals occurs towards a known background dataset, hence, discussion regarding the first type is equally necessary (Abba, Pham, et al., 2020a; Abdulazeez et al., 2023; Ahmad et al., 2021; Farooqui & Mehra, 2018; Jibril et al., 2024; Jibril, Malami, et al., 2023; Mahmoud et al., 2021). Modern data mining employs the use of statistics with other tools, ideas and approaches from computer science, database technology, machine learning, deep learning as well as other classical analytical tools. Data mining and classical statistics differs in many aspects (Abba, Egbueri, Benaafi, Usman, Usman, et al., 2023; Benaafi, Yassin, et al., 2022; Farooqui & Mehra, 2018; Gbadamosi et al., 2023; Ismail et al., 2022).

Machine learning (ML) and artificial intelligence (AI) are considered as promising tools in classification and regression recently, which can be applied in diverse field which include engineering, science, pharmaceutical science, computer science and other health sciences (Alamrouni et al., 2022; Isık et al., 2023; Umar et al., 2023; A. G. Usman, Işik, Abba, et al., 2021; Yassin et al., 2023).

In general, medical doctors as well as psychiatrists prescribe psychoactive drugs. Traditionally, these drugs are categorized into different groups such as; antipsychotics, mood stabilizers, anticonvulsant (or anti-epileptics) and antidepressants. Diazepam (DIA) is one example of these psychoactive drugs. Hence, its significant to determine these drug using suitable instruments and tools (Cunha et al., 2019).

Modeling the chromatographic characteristics of DIA using data mining and AI-based algorithms is essential since high-performance liquid chromatography is one of the most widely used techniques for chemical and pharmacological elucidation. To our understanding, this is the first study in the literature to show how data mining and AI-based models can be used in conjunction with the classical methods such as adaptive-neuro fuzzy inference system (ANFIS), linear regression (LR), and support vector machine for modelling the retention period of DIA using HPLC method.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)
www.abjournals.org

METHODOLOGY

Proposed Models for the simulation

The study employs the use of dual AI algorithms which include SVM, ANFIS and a classical model inform of LR for the simulation of retention behavior of the analyte (DIA) using the HPLC technique. The information used in this research was derived from historical data based on the Jouyban et al research (Jouyban et al., 2009). This work involves the use of two independent variables which include temperature (T) and methanol (MeOH) as the mobile phase for the simulation of retention time (tR) as the dependent variable (Abdullahi et al., 2020). The study's data was split into two categories using 75% for the training phase and 25% for the testing the developed model. The models' performance was evaluated using the k-fold cross-validation approach, which is the most effective method for achieving unbiased model performance prediction with a small data set (A. G. Usman et al., n.d.). The methodology of the work was shown in Figure 1. Prior to the simulation, the data was standardised using the equation 1.

$$X_i = \frac{x_u - x_{min}}{x_{max} - x_{min}} \quad (1)$$

"Xi" denote the amount being normalized, "xu" denote the unnormalized data, "xmin" refers to minimum amount of data, and "x_{max}" functioned as data set's maximum limit..

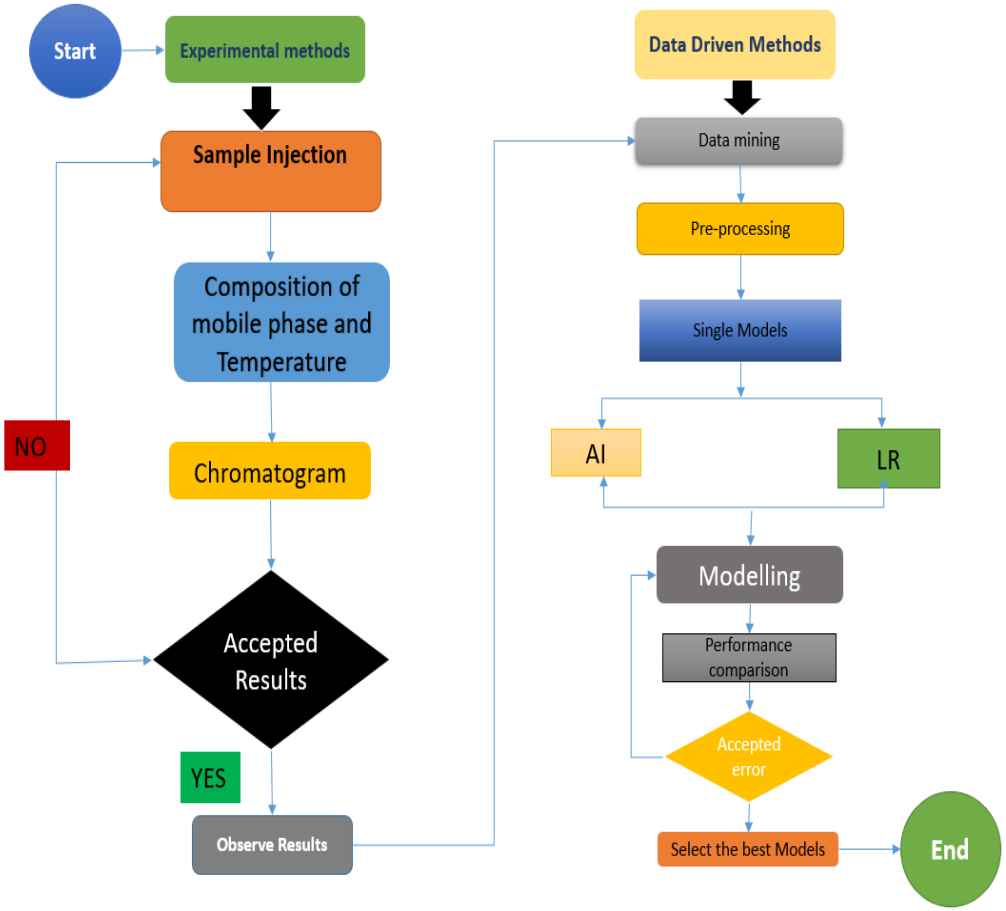

Fig. 1. The adopted flow chart of the methodology

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)

www.abjournals.org

Support Vector Machine (SVM)

Vapnik implemented the support vector machine (SVM) in the year1995 (Vapnik, 1995). SVM is a data-driven algorithm with a learning machine notion that gives a necessary answer to classification, regression, prediction, and pattern recognition problems (Nourani et al., 2018). The SVM's most beneficial function is statistical learning theory and architectural risk minimisation, Reduced error and complexity, as well as increase in the network's performance capacity, are characteristics that differentiate SVM from ANN. The kernel function of the SVM has been applied in various field including prediction of rainfall using the two notable type of SVM which include linear support vector and non-support vector regression (USMAN et al., 2020). Because the SVM comprises layers that include kernel functions weighted on the input variable and weighted sum of kernel output, it can also be conceptualized as SVR (Yaseen et al., 2018). When using the Support Vector Machine, the linear regression was carried out on the data after the non-linear kernel was used to identify a non-linear pattern in the data. The input vector xi serves as a representation of the data set, while di denotes the actual value and N is the total number of data patterns. By mapping positive values for the slack parameters of ξ* and ξ , which is the minimised objective function, it is possible to determine the regression parameters of b and w (Qian et al., 2020).

$$y = f(x) = w\varphi(x_i) + b \tag{2}$$

$$\text{Minimize: } \frac{1}{2}\|w\|^2 + C\left(\sum_i^N (\xi_i + \xi_i^*)\right) \tag{3}$$

$$\text{Subject to: } \begin{cases} w_i\varphi(x_i) + b_i - d_i \le \varepsilon + \xi_i^* \\ d_i - w_i\varphi(x_i) + b_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \end{cases} \quad i=1,2,\ldots,\text{N} \tag{4}$$

Where $\frac{1}{2}\|w\|^2$ are, the vector norm of the weights and *C* refers to the regularized constant. The SVM model's general architecture is depicted in Fig. 2 below. The αi and αi* refers to the parameters of the Lagrange multipliers. Upon determining the problem solution of optimisation, the vector w in Equation (5) can be obtained.

$$w^* = \sum_{i=1}^N (\alpha_i - \alpha_i^*)\varphi(x_i) \tag{5}$$

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^N (\alpha_i - \alpha_i^*)K(x, x_i) + b \tag{6}$$

Where b is the bias term and k (xi, xj) is the kernel function. The most used kernel function is the Radial Basis Function (Gaussian), which is written as:

$$k(x_1, x_2) = \exp(-\gamma\|x_1 - x_2\|^2) \tag{7}$$

where kernel parameter is represented as γ.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)
www.abjournals.org

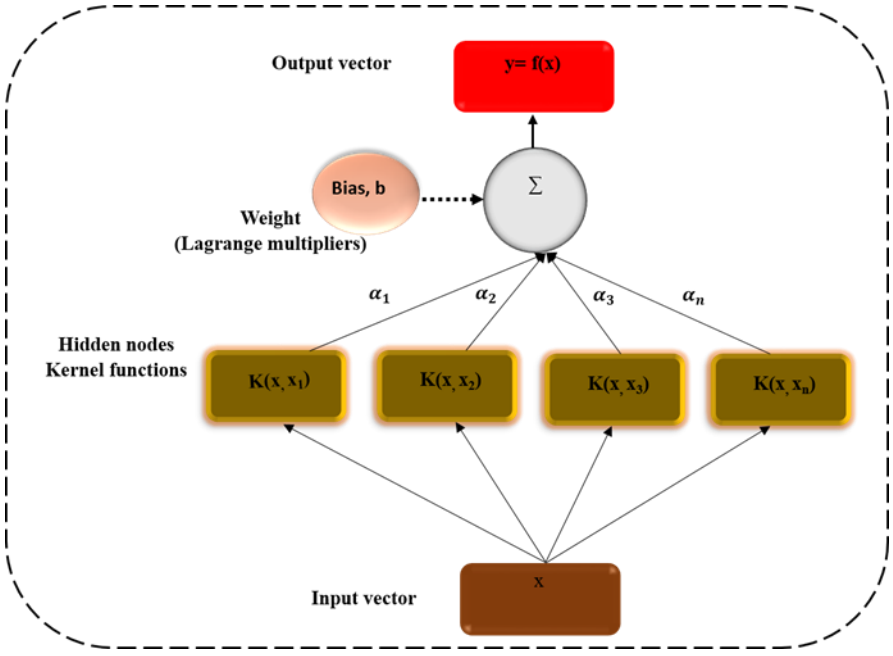

Fig. 2. Architecture of the SVM algorithm utilized in this research

Adaptive - Neuro-Fuzzy Inference System (ANFIS)

ANFIS is a comprehensive model that provides solutions to numerous problems. Feed-forward networks and adaptive multi-layer systems are its important components. Depending on the Takagi-Sugeno form, which consists of input-output variables and a fuzzy rule, the model consists of input variables and a fuzzy rule. The fuzzy network's two fundamental components are the fuzzier and defuzzifier. By employing membership functions, fuzzy logic translates input parameters into fuzzy values (Khalid & Usman, 2021). Nodes serve as membership functions (MFs), enabling the simulation to further unveil the correlation between input and output parameters. The model has trapezoidal triangular, sigmoid, and Gaussian membership functions (Elkiran et al., 2019), (Abba, Usman, et al., 2020).

Based on the assumption that the fuzzy inference system consists of two input variables and one output variable, the rules for a first-order Sugeno fuzzy are as follows.

Rule no 1: assuming $\mu(z_1) is\ C_1\ and\ \mu(z_2) is\ D_1\ therefore\ f_1 = j_1 z_1 + k_1 z_2$ (1)

Rule no 2: assuming $\mu(z_1) is\ C_2\ and\ \mu(z_2) is\ D_2\ therefore\ f_2 = j_2 z_2 + k_2 z_2$ (2)

The variables C_1, D_1, C_2, D_2 inputs MFs of z_1 and z_2, while the output function's variables are j_1, k_1, and j_2, k_2.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)

www.abjournals.org

Linear Regression (LR)

The two types of basic linear regression (LR) are simple regression, which estimates one predictor with one variable, and multiple regression, which estimates many predictors with one variable (multiple regression) (Abba, Usman, Abdulazeez, Lawal, Baig, et al., 2023; Alotaibi et al., 2023; Ghali, Usman, et al., 2020; A. G. Usman, Tanimu, et al., 2023). In this research investigation, the most common type of LR (multilinear regression) was employed (A. G. Usman, Işik, & Abba, 2021). In this kind of analysis, each value of the independent variable will be assigned to a value of the dependent variable. Equation 10 below represents the linear regression used in this research.

$$y = b_0 + b_1 x_1 + b_2 x_2 + ...b_i x_i \quad (10)$$

Where the ith predictor is referred at x_1, b_0 is the constant for regression, and b_i. is the coefficient of ith predictor.

Where x1 is the reference point for the ith predictor, b_i is the ith predictor's coefficient, and b_0 is the regression constant.

Criteria for evaluating the models

In this research, the models were developed taking into account two indicators: the coefficient of determinacy (R^2) and the root mean square error (RMSE).

The R^2 ($-\infty < R^2 < 1$), RMSE ($0 < RMSE < \infty$) are expressed as:

$$R^2 = 1 - \left[\frac{\sum_{i=1}^{N}(\sigma_{com,i} - \sigma_{pre,i})^2}{\sum_{i=1}^{N}(\sigma_{com,i} - \overline{\sigma_{com}})^2}\right] \quad (11)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\sigma_{com,i} - \sigma_{pre,i})^2} \quad (12)$$

Where σ is the strength in compression, in which$\sigma_{pre,i}$,$\sigma_{com,i}$, $\overline{Cd_{pre}}$ and $\overline{\sigma_{com}}$ signifies the predicted and computed σ variables along with their individual N data point means. Furthermore, within the research domain, the recommended models with the highest R^2 and lowest RMSE were nominated for better tR prediction (Ghali, Alhosen, et al., 2020; A. G. Usman, Usanase, et al., 2023; J. Usman, Abba, et al., 2023).

Description of the data set and model validation

The primary objective of effective data mining is to enable development of an algorithm with a set of data from the indicators being used as a framework for accurately forecasting the unknown variables. Given that different constraints, such overfitting, produce training outcomes that do not always correspond to the testing results. Many predictive studies use validation approaches like k-fold cross-validation, holdout, and leaving one out. The main distinction between the k fold cross-validation technique and others is the independence of the validation and training sets (Abba, Pham, et al., 2020b; Benaafi, Tawabini, et al., 2022; Haruna

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

et al., 2021; Selin & Abba, 2020; J. Usman, Salami, et al., 2023). In this research, k-fold cross-validation was used to reduce the issues that might arise due to overfitting. Furthermore, we divided the data used in this study into 25% for testing the developed model and 75% for training the model. However , it's essential to understand that various validation models could also be used for the data set (Abba, Usman, et al., 2020).

RESULTS AND DISCUSSION

The main motivation for this work is the development of different models depending on the regression and machine learning concept with the aid of data mining to predict the chemical properties of DIA. The acquired results are presented in both quantitative and visual form.

Table 1 showed the statistical and correlation analyses of both the independent and dependent variables used in this research.

Table 1: Correlation Analysis

Variables	MeOH	Temperature	DIA (tR)
MeOH	1		
Temperature	0.100768	1	
DIA (tR)	-0.86242	-0.284	1

Table 1 provides the assurance of the correlation co-efficient completed to improve the methodology of the models, where the relationship between the variables is shown by the directional symbols (+ or -). Finding the variables that are most and least related to one another is made easier with the help of the correlation analysis. By providing the input parameter with the highest correlation to the output variable, it also helps with understanding the nature of the data before the modelling process (Usman et al., 2020). Furthermore, Table 1 indicates that there is a solid correlation which happens inversely between the tR of DIA and the mobile phase used in the study inform of MeOH with R-value = -0.86242 and weak inverse correlation between the tR of DIA and temperature having R-value = -0.284.

More also, Table 2 presents the descriptive statistics used in the study. Based on Table 1, the average/mean values of the dependent variable DIA is 4.96% with 2.52 and 16.72 as minimum and maximum values respectively. The minimal skewness levels imply that indeed the experimental results seems to be credible for the modelling process.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Table 2: Descriptive statistics of the study

Parameters	MeOH	Temperature	DIA
Mean	0.81	31.67	4.96
Median	0.80	30.00	4.13
St.D	0.14	5.88	3.54
Kurtosis	0.18	-1.48	9.78
Skewness	-0.45	0.16	2.92
Minimum	0.50	25.00	2.52
Maximum	1.00	40.00	16.72

Performance of the data-driven techniques

MATLAB 9.3 was used to model the data. In terms of the SVM and ANFIS frameworks, the optimum membership function of the ANFIS and cubic SVM was selected based on a trial-and-error technique. The models' predicted performance was evaluated using two separate performance indices. While RMSE was utilised to identify the errors that the algorithms encountered during both the training and testing phases, R^2 was employed to evaluate the fitness between the simulated and actual values.

Table 3: **Performance simulation of the SVM, ANFIS and LR models**

	Training Stage		Testing Stage	
Models	**R^2**	**RMSE.**	**R^2.**	**RMSE.**
SVM.	0.980	0.009	0.970	0.010
ANFIS.	**0.990**	**0.00003**	**0.977**	**0.001**
LR.	0.476	0.223	0.216	0.013

In view of the performance outcomes of the algorithms as shown in Table 2, AI algorithms outperformed the conventional linear regression algorithms in both the training and testing stages by enhancing performance skills up to 51.4 % and 76.1 % for SVM and ANFIS, respectively. This demonstrates the AI-models' superior robustness and efficiency over linear models in terms of capturing the non-linearity and complicated features of a given datasets. The performance skills of the algorithms is equally demonstrated using the scatter plots, to describe the fitness shown by the individual models.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)
www.abjournals.org

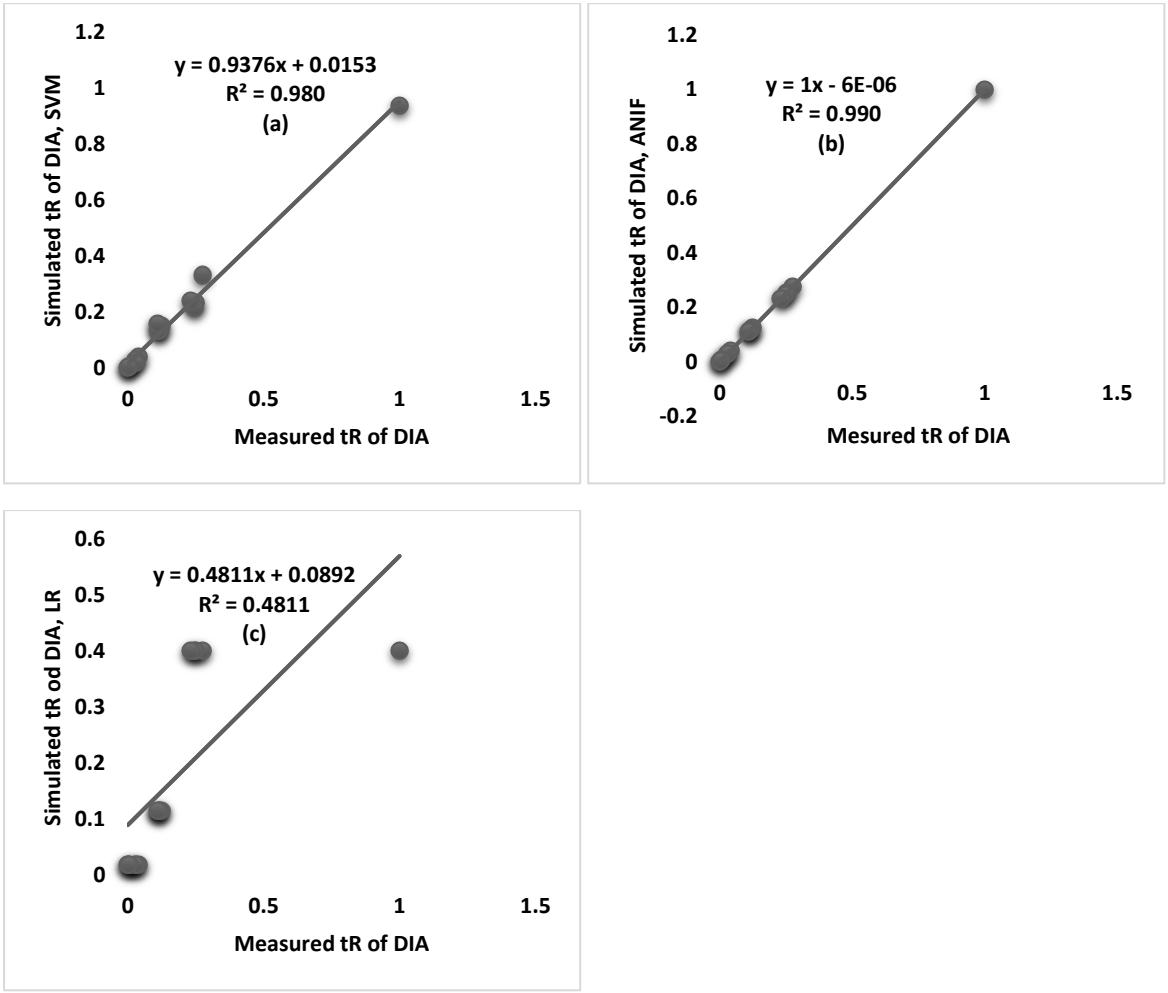

Figure 3: Performance agreement of a) ANFIS b) SVM and c) LR models with the experimental values

Furthermore, the performance of these algorithms is graphically represented using a bar chart depending on their RMSE-values. This corresponds to our findings in Table 3.

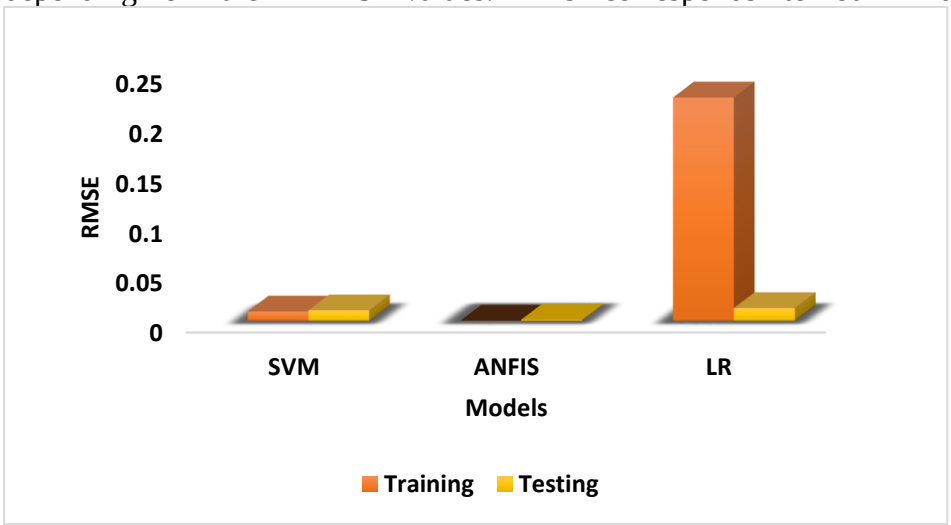

Figure 4: Comparative error performances of the models

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Advanced Journal of Science, Technology and Engineering
ISSN: 2997-5972
Volume 4, Issue 1, 2024 (pp. 119-134)
www.abjournals.org

The radar plot is used to illustrate the models' comparative performance (see Figure 5). The SVM and ANFIS algorithms can be used for mdelling process with several parameters, minimising the output error and eliminating fitting difficulties. The algorithms become a go-to method for solving extremely complex nonlinear difficulties. In the modelling phase, the Radar diagram was used to ascertain the models' comparative performance in terms of the correlation coefficients (R^2). The general performance comparison of the models in modelling the tR of DIA can be described in an ascending order ANFIS>SVM>LR as depicted in figure 5. As a result, the AI-based models outperform the linear model in terms of prediction accuracy. This article demonstrated how the AI-based algorithms are appropriate for the conduct of research in pharmaceutical science, chemical science, computer science, engineering, and life sciences (Gaya et al., 2014; Nourani, 2017; Solgi et al., 2017).

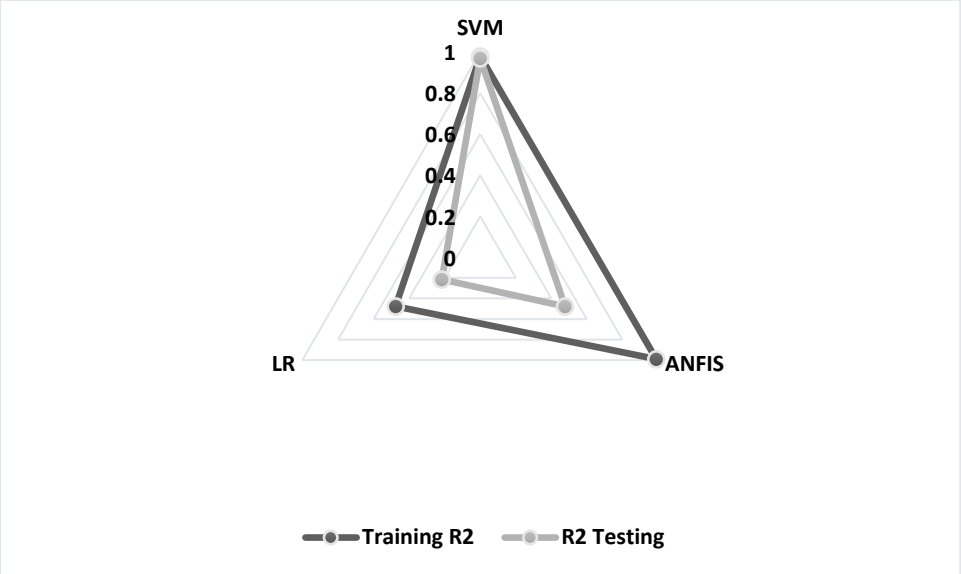

Figure 5: Radar plots for modelling DIA in HPLC technique using ANFIS, SVM and LR in both training and testing phases

The AI models' goodness-of-fit revealed that the algorithms were capable of handling nonlinearity in the DIA retention time variable. Based on the performance evaluation criteria, the DC was found to be 0.990 and 0.977 in both the training and testing phases. Detailed examination of the graphical chart reveals that ANFIS demonstrated overall greater performance over the other algorithms.

CONCLUSION

This current research focus on the modelling of the properties of DIA using HPLC technique. The data involved in the study was extracted from historical data from the published technical literature using the concept of data mining. Three models; SVM, ANFIS and LR models were used in modelling the diazepam agent, which can used in treating psychological disorders. Two performance indices R^2 and RMSE were used in evaluation the fitness and errors depicted by each model respectively. The obtained results demonstrated that AI algorithms are more robust than conventional linear models. However, it is equally recommended to employ other single AI models and emerging optimization techniques to boost the model's performance.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

REFERENCES

Abba, S. I., Abdulkadir, R. A., Sammen, S. S., Usman, A. G., Meshram, S. G., Malik, A., & Shahid, S. (2021). Comparative implementation between neuro-emotional genetic algorithm and novel ensemble computing techniques for modelling dissolved oxygen concentration. *Hydrological Sciences Journal*, *0*(0). https://doi.org/10.1080/02626667.2021.1937179

Abba, S. I., Benaafi, M., Usman, A. G., & Aljundi, I. H. (2022). Inverse groundwater salinization modeling in a sandstone's aquifer using stand-alone models with an improved non-linear ensemble machine learning technique. *Journal of King Saud University - Computer and Information Sciences*, *34*(10), 8162–8175. https://doi.org/10.1016/j.jksuci.2022.08.002

Abba, S. I., Benaafi, M., Usman, A. G., & Aljundi, I. H. (2023). Sandstone groundwater salinization modelling using physicochemical variables in Southern Saudi Arabia : Application of novel data intelligent algorithms. *Ain Shams Engineering Journal*, *14*(3), 101894. https://doi.org/10.1016/j.asej.2022.101894

Abba, S. I., Benaafi, M., Usman, A. G., Ozsahin, D. U., Tawabini, B., & Aljundi, I. H. (2023). Mapping of groundwater salinization and modelling using meta-heuristic algorithms for the coastal aquifer of eastern Saudi Arabia. *Science of The Total Environment*, *858*(July 2022), 159697. https://doi.org/10.1016/j.scitotenv.2022.159697

Abba, S. I., Egbueri, J. C., Benaafi, M., Usman, J., Usman, A. G., & Aljundi, I. H. (2023). Fluoride and nitrate enrichment in coastal aquifers of the Eastern Province, Saudi Arabia: The influencing factors, toxicity, and human health risks. *Chemosphere*, *336*, 139083. https://doi.org/10.1016/j.chemosphere.2023.139083

Abba, S. I., Pham, Q. B., Usman, A. G., Linh, N. T. T., Aliyu, D. S., Nguyen, Q., & Bach, Q. V. (2020a). Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, *33*(October 2019), 101081. https://doi.org/10.1016/j.jwpe.2019.101081

Abba, S. I., Pham, Q. B., Usman, A. G., Linh, N. T. T., Aliyu, D. S., Nguyen, Q., & Bach, Q. V. (2020b). Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, *33*(December 2019), 101081. https://doi.org/10.1016/j.jwpe.2019.101081

Abba, S. I., & Usman, A. G. (2020). *Modeling of Water Treatment Plant Performance using Artificial Neural Network: Case Study Tamburawa Kano-Nigeria*. *September*.

Abba, S. I., Usman, A. G., & Işik, S. (2020). Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach. *Chemometrics and Intelligent Laboratory Systems*, 104007. https://doi.org/10.1016/j.chemolab.2020.104007

Abba, S. I., Usman, J., Abdulazeez, I., Lawal, D. U., Baig, N., Usman, A. G., & Aljundi, I. H. (2023). *Integrated Modeling of Hybrid Nanofiltration / Reverse Osmosis Desalination Plant Using Deep Learning-Based Crow Search Optimization Algorithm*.

Abdulazeez, I., Abba, S. I., Usman, J., Usman, A. G., & Aljundi, I. H. (2023). *Recovery of Brine Resources Through Crown-Passivated Graphene , Silicene , and Boron Nitride Nanosheets Based on Machine-Learning Structural Predictions*. https://doi.org/10.1021/acsanm.3c04421

Abdullahi, H. U., Usman, A. G., & Abba, S. I. (2020). *Modelling the Absorbance of a Bioactive*

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Compound in HPLC Method using Artificial Neural Network and Multilinear Regression Methods. *6*(2), 362–371.

Ahmad, M. H., Usman, A. G., & Abba, S. I. (2021). Comparative performance of extreme learning machine and Hammerstein–Weiner models for modelling the intestinal hyper-motility and secretory inhibitory effects of methanolic leaf extract of Combretumhypopilinum Diels (Combretaceae). *In Silico Pharmacology*, *9*(1). https://doi.org/10.1007/s40203-021-00090-1

Alamrouni, A., Aslanova, F., Mati, S., Maccido, H. S., Jibril, A. A., Usman, A. G., & Abba, S. I. (2022). Multi-Regional Modeling of Cumulative COVID-19 Cases Integrated with Environmental Forest Knowledge Estimation: A Deep Learning Ensemble Approach. *International Journal of Environmental Research and Public Health*, *19*(2), 1–22. https://doi.org/10.3390/ijerph19020738

Alotaibi, B. S., Yahuza, M. S., Ozden, O., Abuhussain, M. A., Dodo, Y. A., Usman, A. G., Usman, J., & Abba, S. I. (2023). *Sustainable Green Building Awareness : A Case Study of Kano Green Construction*.

Bala, K., Etikan, I., Usman, A. G., & Abba, S. I. (2023). Artificial-Intelligence-Based Models Coupled with Correspondence Analysis Visualization on ART—Cases from Gombe State, Nigeria: A Comparative Study. *Life*, *13*(3), 715. https://doi.org/10.3390/life13030715

Benaafi, M., Tawabini, B., Abba, S. I., Humphrey, J. D., AL-Areeq, A. M., Alhulaibi, S. A., Usman, A. G., & Aljundi, I. H. (2022). Integrated Hydrogeological, Hydrochemical, and Isotopic Assessment of Seawater Intrusion into Coastal Aquifers in Al-Qatif Area, Eastern Saudi Arabia. *Molecules*, *27*(20), 1–21. https://doi.org/10.3390/molecules27206841

Benaafi, M., Yassin, M. A., Usman, A. G., & Abba, S. I. (2022). *Neurocomputing Modelling of Hydrochemical and Physical Properties of Groundwater Coupled with Spatial Clustering , GIS , and Statistical Techniques*.

Cunha, D. L., Mendes, M. P., & Marques, M. (2019). Environmental risk assessment of psychoactive drugs in the aquatic environment. *Environmental Science and Pollution Research*, *26*(1), 78–90. https://doi.org/10.1007/s11356-018-3556-z

Elkiran, G., Nourani, V., & Abba, S. I. (2019). Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach. *Journal of Hydrology*, *577*(April), 123962. https://doi.org/10.1016/j.jhydrol.2019.123962

Farooqui, N. A., & Mehra, R. (2018). Design of a data warehouse for medical information system using data mining techniques. *PDGC 2018 - 2018 5th International Conference on Parallel, Distributed and Grid Computing*, 199–203. https://doi.org/10.1109/PDGC.2018.8745864

Gaya, M. S., Abdul Wahab, N., Sam, Y. M., & Samsudin, S. I. (2014). ANFIS Modelling of Carbon and Nitrogen Removal in Domestic Wastewater Treatment Plant. *Jurnal Teknologi*, *67*(5). https://doi.org/10.11113/jt.v67.2839

Gbadamosi, A., Adamu, H., Usman, J., Usman, A. G., Jibril, M. M., Abiodun, B., Lekan, S., Oyedele, L. O., & Abba, S. I. (2023). ScienceDirect New-generation machine learning models as prediction tools for modeling interfacial tension of hydrogen-brine system. *International Journal of Hydrogen Energy*, *xxxx*. https://doi.org/10.1016/j.ijhydene.2023.09.170

Ghali, U. M., Alhosen, M., Degm, A., Alsharksi, A. N., Hoti, Q., & Usman, A. G. (2020). *Development Of Computational Intelligence Algorithms For Modelling The Performance Of Humanin And Its Derivatives In HPLC Optimization Method Development*. *9*(08),

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

110–117.

Ghali, U. M., Usman, A. G., Chellube, Z. M., Degm, M. A. A., Hoti, K., Umar, H., & Abba, S. I. (2020). Advanced chromatographic technique for performance simulation of anti-Alzheimer agent: an ensemble machine learning approach. *SN Applied Sciences*, *2*(11). https://doi.org/10.1007/s42452-020-03690-2

Haruna, S. I., Malami, S. I., Adamu, M., Usman, A. G., Farouk, A., Ali, S. I. A., & Abba, S. I. (2021). Compressive Strength of Self-Compacting Concrete Modified with Rice Husk Ash and Calcium Carbide Waste Modeling: A Feasibility of Emerging Emotional Intelligent Model (EANN) Versus Traditional FFNN. *Arabian Journal for Science and Engineering*, *46*(11), 11207–11222. https://doi.org/10.1007/s13369-021-05715-3

Isık, S., Usman, A. G., & Abba, S. I. (2023). *A chemometrics-based approach for the determination of thymoquinone from Nigella sativa L . (Black Cumin) seeds of different geographical regions using the HPLC technique*. *53*(3), 320–328. https://doi.org/10.26650/IstanbulJPharm.2023.1181298

Ismail, S., Usman, R. A. A. A. G., & Abba, S. I. (2022). Development of chemometrics - based neurocomputing paradigm for simulation of manganese extraction using solid - phase tea waste. *Modeling Earth Systems and Environment*. https://doi.org/10.1007/s40808-022-01369-8

Jibril, M. M., Idris, S., Hauwa, M., Jibrin, B., Jibrin, U., Adamu, M., Usman, D. A. G., Dilber, A. G. I., Ozsahin, U., & Karami, Z. (2024). New random intelligent chemometric techniques for sustainable geopolymer concrete : low - energy and carbon - footprint initiatives. *Asian Journal of Civil Engineering*, *25*(2), 2287–2305. https://doi.org/10.1007/s42107-023-00908-7

Jibril, M. M., Malami, S. I., Muhammad, U. J., Bashir, A., Usman, A. G., Salami, B. A., Rotimi, A., Ibrahim, A. G., & Abba, S. I. (2023). High strength concrete compressive strength prediction using an evolutionary computational intelligence algorithm. *Asian Journal of Civil Engineering*, *0123456789*. https://doi.org/10.1007/s42107-023-00746-7

Jibril, M. M., Zayyan, M. A., Idris, S., Usman, A. G., Salami, B. A., Rotimi, A., & Abba, S. I. (2023). Applications in Engineering Science Implementation of nonlinear computing models and classical regression for predicting compressive strength of high-performance concrete. *Applications in Engineering Science*, *15*(April), 100133. https://doi.org/10.1016/j.apples.2023.100133

Jouyban, A., Soltanpour, S., Acree, W. E., Thomas, D., Agrafiotou, P., & Pappa-Louisi, A. (2009). Modeling the effects of different mobile phase compositions and temperatures on the retention of various analytes in HPLC. *Journal of Separation Science*, *32*(22), 3898–3905. https://doi.org/10.1002/jssc.200900389

Khalid, G. M., & Usman, A. G. (2021). *Application of data-intelligence algorithms for modeling the compaction performance of new pharmaceutical excipients*.

Madaki, Z., Abacioglu, N., Usman, A. G., Taner, N., Sehirli, A. O., & Abba, S. I. (2022). Novel Hybridized Computational Paradigms Integrated with Five Stand-Alone Algorithms for Clinical Prediction of HCV Status among Patients: A Data-Driven Technique. *Life*, *13*(1), 79. https://doi.org/10.3390/life13010079

Mahmoud, K., Bebiş, H., Usman, A. G., Salihu, A. N., Gaya, M. S., Dalhat, U. F., Abdulkadir, R. A., Jibril, M. B., & Abba, S. I. (2021). Prediction of the effects of environmental factors towards COVID-19 outbreak using AI-based models. *IAES International Journal of Artificial Intelligence*, *10*(1), 35–42. https://doi.org/10.11591/ijai.v10.i1.pp35-42

Nourani, V. (2017). An Emotional ANN (EANN) approach to modeling rainfall-runoff process. *Journal of Hydrology*, *544*, 267–277.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

https://doi.org/10.1016/j.jhydrol.2016.11.033

Nourani, V., Elkiran, G., & Abba, S. I. (2018). Wastewater treatment plant performance analysis using artificial intelligence - An ensemble approach. *Water Science and Technology*, *78*(10), 2064–2076. https://doi.org/10.2166/wst.2018.477

Pham, Q. B., Abba, S. I., Usman, A. G., Linh, N. T. T., Gupta, V., Malik, A., Costache, R., Vo, N. D., & Tri, D. Q. (2019). Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall. *Water Resources Management*. https://doi.org/10.1007/s11269-019-02408-3

Qian, N., Wang, X., Fu, Y., Zhao, Z., Xu, J., & Chen, J. (2020). Predicting heat transfer of oscillating heat pipes for machining processes based on extreme gradient boosting algorithm. *Applied Thermal Engineering*, *164*(July 2019), 114521. https://doi.org/10.1016/j.applthermaleng.2019.114521

Rezaei-Darzi, E., Farzadfar, F., Hashemi-Meshkini, A., Navidi, I., Mahmoudi, M., Varmaghani, M., Mehdipour, P., Alamdari, M. S., Tayefi, B., Naderimagham, S., Soleymani, F., Mesdaghinia, A., Delavari, A., & Mohammad, K. (2014). Comparison of two data mining techniques in labeling diagnosis to Iranian Pharmacy Claim Dataset: Artificial Neural Network (ANN) versus decision tree model. *Archives of Iranian Medicine*, *17*(12), 837–843. https://doi.org/0141712/AIM.0010

Rong, W., Guo, S., Ding, K., Yuan, Z., Li, Q., & Bi, K. (2018). Integrated strategy based on high-resolution mass spectrometry coupled with multiple data mining techniques for the metabolic profiling of Xanthoceras sorbifolia Bunge husks in rat plasma, urine, and feces. *Journal of Separation Science*, *41*(13), 2846–2853. https://doi.org/10.1002/jssc.201800012

Selin, A. G. U., & Abba, I. S. I. (2020). A Novel Multi - model Data - Driven Ensemble Technique for the Prediction of Retention Factor in HPLC Method Development. *Chromatographia*, *0123456789*. https://doi.org/10.1007/s10337-020-03912-0

Sohrabi, B., Raeesi Vanani, I., Nikaein, N., & Kakavand, S. (2019). A predictive analytics of physicians prescription and pharmacies sales correlation using data mining. *International Journal of Pharmaceutical and Healthcare Marketing*, *13*(3), 346–363. https://doi.org/10.1108/IJPHM-11-2017-0066

Solgi, A., Pourhaghi, A., Bahmani, R., & Zarei, H. (2017). Improving SVR and ANFIS performance using wavelet transform and PCA algorithm for modeling and predicting biochemical oxygen demand (BOD). *Ecohydrology and Hydrobiology*, *17*(2), 164–175. https://doi.org/10.1016/j.ecohyd.2017.02.002

Umar, H., Rizaner, N., Usman, A. G., Aliyu, M. R., Adun, H., Ghali, U. M., Ozsahin, D. U., & Abba, S. I. (2023). *Prediction of Cell Migration in MDA-MB 231 and MCF-7 Human Breast Cancer Cells Treated with Albizia Lebbeck Methanolic Extract Using Multilinear Regression and Artificial Intelligence-Based Models*.

Usman, A. G., Ahmad, M. H., Danraka, N., & Abba, S. I. (2021). The effect of ethanolic leaves extract of Hymenodictyon floribundun on inflammatory biomarkers : a data - driven approach. *Bulletin of the National Research Centre*. https://doi.org/10.1186/s42269-021-00586-y

Usman, A. G., Ghali, U. M., Degm, M. A. A., Muhammad, S. M., Hincal, E., Kurya, A. U., Işik, S., Hoti, Q., & Abba, S. I. (2022). Simulation of liver function enzymes as determinants of thyroidism: a novel ensemble machine learning approach. *Bulletin of the National Research Centre*, *46*(1). https://doi.org/10.1186/s42269-022-00756-6

Usman, A. G., Işik, S., & Abba, S. I. (n.d.). *A Novel Multi-model Data-Driven Ensemble Technique for the Prediction of Retention Factor in HPLC Method Development*.

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Usman, A. G., Işik, S., & Abba, S. I. (2021). Hybrid data-intelligence algorithms for the simulation of thymoquinone in HPLC method development. *Journal of the Iranian Chemical Society*, *0123456789*. https://doi.org/10.1007/s13738-020-02124-5

Usman, A. G., IŞIK, S., & Abba, S. I. (2022). Qualitative prediction of Thymoquinone in the high-performance liquid chromatography optimization method development using artificial intelligence models coupled with ensemble machine learning. *Separation Science Plus*, *June*, 1–9. https://doi.org/10.1002/sscp.202200071

Usman, A. G., Işik, S., Abba, S. I., & Meriçli, F. (2021). Chemometrics-based models hyphenated with ensemble machine learning for retention time simulation of isoquercitrin in Coriander sativum L. using high-performance liquid chromatography. *Journal of Separation Science*, *December 2020*, 1–7. https://doi.org/10.1002/jssc.202000890

USMAN, A. G., IŞIK, S., ABBA, S. I., & MERİÇLİ, F. (2020). Artificial intelligence–based models for the qualitative and quantitative prediction of a phytochemical compound using HPLC method. *Turkish Journal of Chemistry*, *44*(5), 1339–1351. https://doi.org/10.3906/kim-2003-6

Usman, A. G., Tanimu, A., Abba, S. I., Isik, S., Aitani, A., & Alasiri, H. (2023). *Feasibility of the Optimal Design of AI-Based Models Integrated with Ensemble Machine Learning Paradigms for Modeling the Yields of Light Olefins in Crude-to-Chemical Conversions*. https://doi.org/10.1021/acsomega.3c05227

Usman, A. G., Usanase, N., Abba, S. I., Ozsahin, I., Uzun, B., Yassin, M. A., Rahman, S. M., & Ozsahin, D. U. (2023). Environmental modelling of CO concentration using AI-based approach supported with filters feature extraction_ A direct and inverse chemometrics-based simulation. *Sustainable Chemistry for the Environment*, *2*(February), 100011. https://doi.org/10.1016/j.scenv.2023.100011

Usman, J., Abba, S. I., Babatunde, N., Hafiz, M., Othman, D., & Aljundi, I. H. (2023). Genetic neuro-computing model for insights on membrane performance in oily wastewater treatment : An integrated experimental approach. *Chemical Engineering Research and Design*, *199*, 33–48. https://doi.org/10.1016/j.cherd.2023.09.027

Usman, J., Salami, B. A., Gbadamosi, A., Adamu, H., Usman, A. G., Benaafi, M., Abba, S. I., Hafiz, M., Othman, D., & Aljundi, I. H. (2023). Chemosphere Intelligent optimization for modelling superhydrophobic ceramic membrane oil flux and oil-water separation efficiency : Evidence from wastewater treatment and experimental laboratory. *Chemosphere*, *331*(March), 138726. https://doi.org/10.1016/j.chemosphere.2023.138726

Uzun Ozsahin, D., Balcioglu, O., Usman, A. G., Ikechukwu Emegano, D., Uzun, B., Abba, S. I., Ozsahin, I., Yagdi, T., & Engin, C. (2022). Clinical Modelling of RVHF Using Pre-Operative Variables: A Direct and Inverse Feature Extraction Technique. *Diagnostics*, *12*(12), 3061. https://doi.org/10.3390/diagnostics12123061

Uzun Ozsahin, D. U., Precious Onakpojeruo, E. P., Bartholomew Duwa, B., Usman, A. G., Isah Abba, S. I., & Uzun, B. (2023). COVID-19 Prediction Using Black-Box Based Pearson Correlation Approach. *Diagnostics*, *13*(7), 1264. https://doi.org/10.3390/diagnostics13071264

Vapnik, V. N. (1995). The nature of statistical learning theory. *Springer, New York*.

Yaseen, Z. M., Deo, R. C., Hilal, A., Abd, A. M., Bueno, L. C., Salcedo-Sanz, S., & Nehdi, M. L. (2018). Predicting compressive strength of lightweight foamed concrete using extreme learning machine model. *Advances in Engineering Software*, *115*(April 2017), 112–125. https://doi.org/10.1016/j.advengsoft.2017.09.004

Yassin, M. A., Tawabini, B., Al-shaibani, A., Adetoro, J. A., Benaafi, M., Al-areeq, A. M.,

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL

Usman, A. G., & Abba, S. I. (2022). *Geochemical and Spatial Distribution of Topsoil HMs Coupled with Modeling of Cr Using Chemometrics Intelligent Techniques : Case Study from Dammam Area , Saudi Arabia*. 1–19.

Yassin, M. A., Usman, A. G., Abba, S. I., Uzun, D., & Aljundi, I. H. (2023). Results in Engineering Intelligent learning algorithms integrated with feature engineering for sustainable groundwater salinization modelling : Eastern Province of Saudi Arabia. *Results in Engineering*, *20*(August), 101434. https://doi.org/10.1016/j.rineng.2023.101434

Article DOI: 10.52589/AJSTE-XTWJGQQL
DOI URL: https://doi.org/10.52589/AJSTE-XTWJGQQL