İstanbul Journal of Pharmacy
Istanbul J Pharm2024, 54 (2): 133–143
DOI: 10.26650/IstanbulJPharm.2024.1225463

Original Article

Effects of chromatographic conditions on retention behaviour of different psychoactive agents in high-performance liquid chromatography: A machine-learning-based approach

Abdullahi Garba Usman[1], Emine Erdağ[2], Selin Işık[1]

[1]Near East University, Faculty of Pharmacy, Department of Analytical Chemistry, Nicosia, Cyprus
[2]Near East University, Faculty of Pharmacy, Department of Pharmaceutical Chemistry, Nicosia, Cyprus

ABSTRACT

Background and Aims: High-pressure liquid chromatography (HPLC) data on the effects of various chromatographic conditions on the retention behaviour of three different psychotropic drugs; clonazepam, diazepam, and oxazepam) were considered for simulation using a machine learning approach.

Methods: For the simulation of selected psychoactive compounds using HPLC, different machine learning techniques were used in this study: adaptive neuro-fuzzy inference system, multilayer perceptron, Hammerstein-Weiner model, and a traditional linear model in the form of stepwise linear regression. Four evaluation criteria were used to assess the effectiveness of the models: coefficient of determination, root mean squared error, mean squared error, and correlation coefficient.

Results: The results show that machine learning approaches, especially multilayer perceptions, are more reliable than classical linear models with an average coefficient of determination value of 0.98 in both calibration and validation phases.

Conclusion: The performance results also demonstrate that these models can be improved using additional approaches, such as hybrid models, ensemble machine learning, evolving algorithms, and optimisation techniques.

Keywords: Machine learning, clonazepam, diazepam, oxazepam, validation, evaluation metrics

INTRODUCTION

Psychiatrists and general practitioners typically prescribe psychoactive medications. These medications are categorised into several classes, including antipsychotics, mood stabilisers, anticonvulsants (also known as anti-epileptics), and antidepressants (Jouyban et al., 2009). They are also substances that can depress brain activity, such as alcohol, benzodiazepines (e.g., Alprazolam, Diazepam), and barbiturates. These drugs may promote relaxation, sedation, and euphoria. However, they can also cause side effects like impaired coordination, slurred speech, and respiratory depression, when taken at high doses (Jouyban et al., 2009). Previously, antidepressants were only used to treat depression, anti-epileptic drugs were only used to treat epilepsy, antipsychotic drugs were only used to treat acute mania and schizophrenia, and mood stabilisers were only used to treat bipolar disorders. The analytical techniques employed in the clarification and identification of various psychoactive substances, particularly benzodiazepines like clonazepam (CLO), diazepam (DIA), and oxazepam (OXA), have been covered in previous research (Cunha, Mendes, & Marques, 2019). It is crucial to develop a practical analytical method that can be used to identify these compounds due to problems like adulteration, fraud, and abuse of pharmaceutical medications. Technical research articles have therefore used a variety of approaches to determine the qualitative and quantitative features of these psychoactive compounds, including spectroscopy, chromatography, and chemometrics. It is necessary to optimise the chromatographic conditions to achieve the global optimum conditions that can be used in either qualitative determination, quantitative determination, or hybrid determination of the analyte of interest to develop protocols for determining any analyte using chromatographic methods like HPLC (Usman, Isik, & Abba, 2020). Since HPLC is one of the most frequently used analytical methods for pharmaceutical, food, and chemical analysis, simulation of the retention behaviour is one of the crucial topics in the chemometrics approach. Analytes are typically resolved using HPLC based on their distribution between mobile and stationary phases, which is presented in the form of retention time (tR) or occasionally informed of retention factor (k), especially if the retention of the mobile phase, known as the dead

Corresponding Author: Emine Erdağ **E-mail:** emine.erdag@neu.edu.tr
Submitted: 29.12.2022 • **Revision Requested:** 11.04.2024 • **Last Revision Received:** 18.04.2024 • **Accepted:** 27.06.2024

time, is taken into consideration (Usman, Isik, & Abba, 2020). In HPLC, tR is an important parameter for both qualitative and quantitative analysis. The values of tR increased when the analyte interacted strongly with the stationary phase and decreased when the analyte interacted strongly with the mobile phase. As a result, tR is linked to three different parameters. These parameters can be summarised as the structure of the stationary phase, the chemical properties of the analyte, and the physicochemical behaviour of the mobile phase (Usman, Isik, & Abba, 2021). One commonly used technique to enhance the performance of HPLC systems involves increasing solvent strength by adding an organic modifier to the aqueous medium. This increases the resolution of two or more subsequent peaks. Another parameter that influences tR is also taken into account, and that is the column temperature. Traditionally, an expensive and time-consuming trial-and-error method is used to adjust the mobile phase composition, column temperature, and other chromatographic parameters (Ghali et al., 2020). Chemometric techniques can be used to predict chromatographic variables under ideal conditions, such as tR and resolution (Silva et al., 2020).

Currently, several chemical and pharmaceutical branches use chemometrics-based methodologies that were established through regression analysis, classification analysis, and molecular structure descriptors (Erdag, Haskologlu, Mercan, Abacioglu, & Sehirli, 2023; Sultanoglu, Erdag, & Ozverel, 2023; Tomić et al., 2020). However, as these approaches typically use linear regression techniques, such as stepwise linear regression (SWLR), single linear regression (SLR), and multilinear regression (MLR), it is often difficult to make physical sense of these models (D'Archivio, 2019). These descriptors are often multidimensionally complex; thus, non-linear approaches such as artificial intelligence (AI) technologies are the only way to address them (Rodríguez-Díaz & Sánchez-León, 2019). Conventional linear, static analysis, and other statistical techniques are generally used in tR prediction. Statistical/chemometric and experimental design methodologies should be combined to evaluate the properties of various analytes and improve their physical behaviours (Tayyebi, Hajjar, & Soltanali, 2019). Regression analysis seems straightforward and uncomplicated; however, it has some disadvantages when dealing with several independent variables and reduces accuracy. The accuracy of predicted parameters can be improved using multiple regression analysis when creating a model (Abba et al., 2020). Recent technical research has used artificial intelligence and optimisation techniques to address complicated issues in chromatography and other instrumental methods of investigation. For instance, previous studies have described the use of machine learning to predict the retention characteristics of distinct small compounds under various HPLC settings (Osipenko et al., 2020). With an average absolute error of 46 s, the proposed strategy outperformed previously used learning algorithms. A recent study reported the use of artificial intelligence and ensemble machine learning to predict the presence of methyclothiazide and amiloride drugs using the HPLC technique (Usman, Isik, & Abba, 2020). The proposed strategy for estimating the k-values of agents demonstrated robust performance. Furthermore, the use of hybrid models to predict the retention behaviour of thymoquinone using HPLC has been recently demonstrated (Usman, Isik, & Abba, 2020). The method exhibits the hybrid technique's capacity to more effectively simulate both the linear and nonlinear characteristics of the data.

In addition, a method for selecting an appropriate solvent gradient for separating preparative peptides using machine learning was developed (Samuelsson, Eiriksson, Åsberg, Thorsteinsdóttir, & Fornstedt, 2019). The procedure demonstrated that the machine learning approach was effective in showing how chromatographic settings affect retention behaviour. Artificial neural networks (ANN) were used in a previous study to clarify different types of amino acids using the HPLC technique (D'Archivio, 2019). In this study, multilayer perceptions (MLP), adaptive neuro-fuzzy inference systems (NF), Hammerstein-Weiner (HW), and SWLR were used to simulate the effects of selected psychoactive agents using HPLC for the first time. The main goal of this study was to use various machine learning approaches to investigate the physicochemical effects of two chromatographic conditions; column temperature and mobile phase, on the retention behaviour of three psychotropic substances.

MATERIALS AND METHODS

Machine learning techniques

Any data-driven technique must consider scientific principles and data knowledge (Elkiran, Nourani & Abba, 2019). Temperature and mobile phase were used as independent variables to model the tR of three psychoactive substances (CLO, DIA, and OXA). This research used machine learning approaches, such as MLP, NF, HW, and SWLR models. MATLAB 9.3 was used to perform the simulation. Then, the data were divided into 60% calibration and 40% verification. The best way to obtain unbiased estimates of model performance for a small dataset is to use the k-fold cross-validation technique, which is used to test the performance of models for validation purposes. A flowchart of the model process is shown in Figure 1.

The dataset was standardised to range from zero to one (standard scale) to increase the data integrity and reduce redundancy using Equation (**1**).

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

"X_i" served as a normalised quantity, "x_u" served as an unnormalised quantity, "x_{min}" served as the minimum, and "x_{max}" served as the maximum quantity of the dataset.

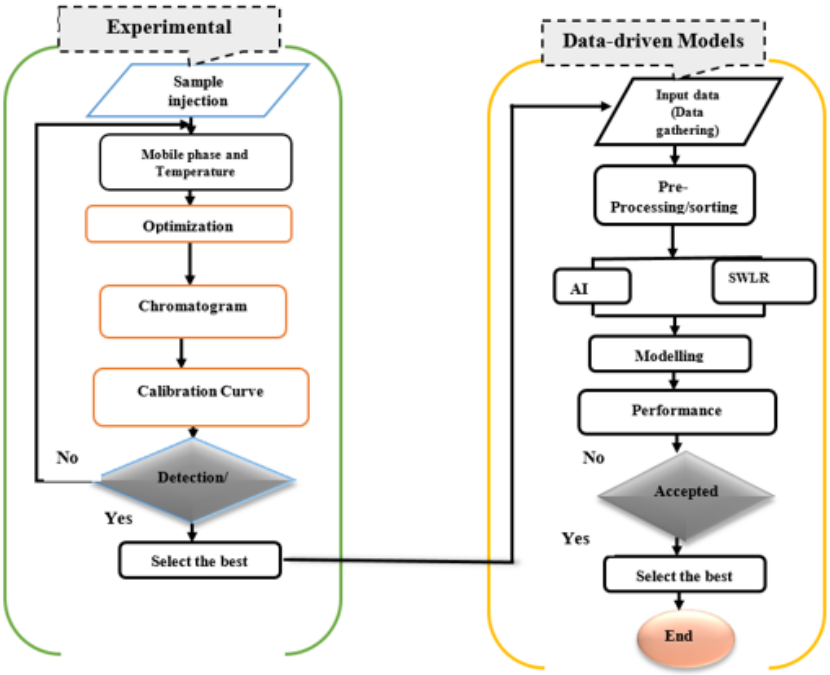

Figure 1. The flow chart of the proposed methodology.

Neuro-fuzzy algorithm

Neuro-fuzzy, a data-intelligence algorithm, mixes fuzzy logic (FL) and neural network learning to handle ambiguous phenomena in collaborative operations. NF is employed as an assessment tool in the field because of its capacity to resemble real functions. On the other hand, the surgeon approach has wider application areas (Choubin, Khalighi-Sigaroodi, Malekian & Kişi, 2016). The defuzzifier and defuzzifier are the basic building blocks of a fuzzy database system. Fuzzy logic involves the use of membership functions to transform input data into fuzzy values ranging from 0 to 1. Nodes that serve as membership functions (MFs) make the input-output relationship modelling possible. Numerous shapes and sizes of the main functions exist, including triangular, sigmoid, gaussian, and trapezoidal (Shojaeimehr & Rahimpour, 2018).

Considering that a first-order Sugeno fuzzy has two inputs (x_1 and x_2) and one output (f), the following principles apply, as shown in Equations (**2**) and Equation (**3**).

$$\text{Law 1 : If } \mu(x_1) is\ A_1\ and\ \mu(x_2) is\ B_1\ then\ f_1 = p_1x_1 + q_1x_2 \quad (2)$$

$$\text{Law 2 : If } \mu(x_1) is\ A_2\ and\ \mu(x_2) is\ B_2\ then\ f_2 = p_2x_2 + q_2x_2 \quad (3)$$

Parameters A_1, B_1, A_2, and B_2 are the main functions for the x_1 and x_2 inputs. The parameters of the outlet function were p_1, q_1, r_1, and p_2, q_2, r_2. The NF structure and formulation were based on a five-layer neural network architecture. Extensive studies have been conducted on the NF model in the literature (Karimi, Kisi, Shiri & Makarynskyy, 2013; Abba, Hadi & Abdullahi, 2017).

Hammerstein-Weiner model

A nonlinear block precedes or opposes a linear dynamic structure in the Hammerstein-Wiener model. The HW was created as a black-box model to identify nonlinear systems. The HW structure is composed of parallel and series-connected nonlinear dynamic and static blocks. Unlike other typical ANNs, the HW model block has been identified as an excellent example that is directly related to linear and nonlinear systems (Abba et al., 2020). The HW model also includes a simple and flexible approach to selecting parametric values for nonlinear models, effectively capturing the physical details of the system properties (Pham et al., 2019).

Multilayer perceptron neural network

A multilayer perception neural network is one of the most prevalent types of ANNs that can handle a nonlinear environment. Compared with other classes of ANNs, researchers consider

this universal approximator (Ghorbani, Deo, Yaseen, Kashani & M0hammadi, 2017). The structure of the MLP consists of an input layer, a hidden layer, and an output layer, as in other well-known ANNs (Kim & Singh, 2014). The signal is translated and sent using weights and biases from the input layer to the output layer through a series of mathematical processes. The Levenberg-Marquardt algorithm was employed as the learning algorithm to reduce the error between the observed and predicted values. The set training procedures were repeated until the intended results were obtained. MLP has an input, one or more hidden layers, and output layers in a structure similar to an ANN.

$$y_i = \sum_{j=1}^{N} w_{ji} x_j + w_{i0} \tag{4}$$

In Equation (**4**), the total number of nodes within the top layer of a node is denoted by N. w_{ji} was the weight between nodes *i* and *j* in the upper layer; x_j was the output derived from node *j*; w_{i0} was the bias in the node *i* and y_i was the input signal of node *i* that crosses through the transfer function.

Evaluation metrics and machine learning methods

In general, linear regression (LR) is one of the most commonly used computational techniques for modelling various input and output variables is linear regression (LR). When choosing the optimal set of parameters for the best prediction accuracy of the output variable, it is important to keep in mind that there is a link between single and multiple variables. According to several modellers, systematic regression is an advancement in selection that makes use of the optimal input dataset by removing or including variables while considering the residual sum of the squares into account (Abba et al., 2020). The SWLR model adheres to the variables' systematic shifts by assessing their effects. To reduce the impact of any variable, each variable that does not support and fulfil the mechanism of the model is eliminated one by one. MLR can be used to explain the concept of the SWLR model (Lee, Han, Lee, & Yoon, 2017).

Any data-driven approach's output accuracy is often evaluated using a variety of metrics based on a comparison of calculated and anticipated values (Usman, Isik, Abba & Meriçli, 2021). To determine the determination coefficient (DC) (Equation (**5**)), correlation coefficient (CC) (Equation (**6**)), and evaluation of the two models (Chandwani, Vyas, Agrawal, & Sharma, 2015), we used two statistical errors, namely root-mean-squared error (RMSE) and mean-squared error (MSE), as shown in Equations (**7**) and (**8**).

$$DC = 1 - \frac{\Sigma_{j=1}^{N}[(Y)_{obs,j} - (Y)_{com,j}]^2}{\Sigma_{j=1}^{N}[(Y)_{obs,j} - \overline{(Y)}_{obs,j}]^2} \tag{5}$$

$$CC = \frac{\Sigma_{i=1}^{N}(Y_{obs} - \overline{Y}_{obs})(Y_{com} - \overline{Y}_{com})}{\sqrt{\Sigma_{i=1}^{N}(Y_{obs} - \overline{Y}_{obs})^2}\Sigma_{i=1}^{N}(Y_{com} - \overline{Y}_{com})^2} \tag{6}$$

$$MSE = \sqrt{\frac{\Sigma_{i=1}^{N}(Y_{obsi} - Y_{comi})^2}{N}} \tag{7}$$

$$MSE = \frac{1}{N}\Sigma_{i=1}^{N}(Y_{obsi} - Y_{comi})^2 \tag{8}$$

Here, N, Y_{obsi}, $\overline{Y}$ *and* Y_{comi} were data number, data observed, average value of the data observed, and computed values, respectively.

Data set and model validation

The data used in the present study were collected from a previous experimental study (Jouyban et al., 2009). The main objective of successful data-driven techniques is to ensure that the model provides a set of data from indicators that can be used as a basis for reliably predicting unknown variables. Considering restrictions like overfitting, the satisfactory training performance may not always match the testing performance. The most important advantage of the k-fold cross-validation procedure is that the validation set and training sets are independent in each round (Samuelsson et al., 2019). Therefore, taking into account k-fold cross-validation, we divided the data in this study into 60% for the calibration (training) phase and 40% for validation. Thus, we have demonstrated the importance of validating the dataset using additional techniques, as reported in a similar study (Abba, Usman, & Isik, 2020).

RESULTS AND DISCUSSION

Retention time is critically important for detecting fraud and misuse in complex matrix media such as food, drinks, drugs, and environmental samples. First, when substances are analysed using chromatographic techniques like HPLC, the retention time serves as a unique fingerprint for the compounds. Deviations from the expected retention times may indicate fraud, such as the addition of foreign chemicals as adulterants or illegal drugs. Analysts can detect potential fraud or adulteration by comparing the retention times of suspicious samples with those of genuine reference standards (Zhang et al., 2024).

Second, in the pharmaceutical industry, retention time is essential for verifying the authenticity of medication compositions. Genuine pharmaceuticals may exhibit different retention properties compared to counterfeit drugs due to the use of substandard or inactive ingredients. Analysts can identify discrepancies that may counterfeit medications by examining the

retention times of excipients and active pharmaceutical ingredients (APIs) (Fine, Mann, & Aggarwal, 2024).

Third, retention time is utilised in forensic toxicology and drug testing to detect and quantify drugs of abuse in biological samples, such as blood, urine, and hair. Variations in the retention time may indicate adulterants or illicit drug analogues, which are substances used to mask drugs in standard screening procedures. Forensic laboratories can enhance the detection of new chemicals or adulterants in biological specimens by developing retention time libraries for well-known drugs and their metabolites (Zhang et al., 2024).

The way machine learning works is based on learning interactions and identifying complex nonlinear patterns in empirical data. These techniques are also effective in the absence of predetermined regression equations with deliberate choices. A regression problem with temperature and mobile phase as independent variables was used to predict and simulate the retention characteristics of psychotropic substances. In this study, the experimental data from previous analyses were applied under various chromatographic conditions. Before modelling, the data were preprocessed with data normalisation and partitioning.

Although HPLC is a well-established technique for determining retention times, it can be time-consuming and resource-intensive. Machine learning models built on existing HPLC data can potentially predict benzodiazepine retention times more quickly and efficiently. This approach might be particularly useful in scenarios requiring rapid analysis or high-throughput screening (Osipenko et al., 2020). However, conducting HPLC experiments, especially with numerous samples or compounds, can be expensive due to the need for specialised equipment, reagents, and skilled personnel. Predictive models can help reduce these costs by minimising the number of samples required for experimental analysis, allowing researchers to focus on the most promising candidates (Sahu et al., 2018).

Furthermore, machine learning models can reveal complex relationships between molecular structures and retention behaviours that are not evident through traditional analytical methods. By analysing a large dataset of benzodiazepines with known retention times, these models can identify structural features or descriptors that significantly influence retention behaviour, offering insights into the molecular mechanisms involved (Usman, Isik, & Abba, 2022).

In analytical laboratories, predictive models can also serve as valuable tools for quality assurance and control. Researchers can quickly identify potential deviations or anomalies by comparing predicted retention times with experimental data, ensuring the accuracy and reliability of analytical results (Isik, Usman, & Abba, 2023). Additionally, machine learning models based on retention time data can be integrated into virtual screening pipelines to prioritise drugs for further experimental validation. By predicting retention times for large libraries of virtual compounds, researchers can accelerate the drug development process by prioritising candidates with suitable retention properties for synthesis or biological testing.

Moreover, while retention time data are crucial for qualitative analysis, particularly in chromatographic separations, analysts must consider the potential impact of matrix effects on retention time predictions. They should take steps to mitigate these effects through method optimisation, sample preparation, and the use of complementary analytical techniques, such as mass spectrometry (MS), to confirm compound identity based on molecular weight, fragmentation patterns, and other structural information. Additionally, calibration curves or standard addition methods can help address matrix effects by accounting for variations in sample composition, ensuring accurate compound identification in complex samples (Gibbs, 2014).

Equation (**1**) was used to standardise the input and target data prior to using the machine learning techniques. The purpose of the normalisation process is to eliminate redundant information and create a uniform scale for the data. The data integrity was improved using the traditional normalisation procedure. The descriptive data and basi details were reported in a previous study (Hadi et al., 2019). The dependent variables CLO, DIA, and OXA have mean and mean values of 3.45%, 4.96%, and 3.75%, respectively. According to the low skewness values, the experimental data were considered reliable for analysis.

The correlation matrix of past experimental data is linearly presented in Figure 2 to highlight the key input variables. Furthermore, the matrix exhibited a weak inverse relationship with temperature and a strong inverse relationship between the mobile phase and the tR of CLO, DIA, and OXA, respectively. The application of traditional modelling techniques to such complex interactions is trivial, as evidenced by the weakness of correlation values; thus, it is imperative to develop more robust and flexible tools.

The optimal architectures for the MLP, HW, SWLR, and NF models was optimised and selected using trial and error. RMSE and MSE were used to describe the error depicted by the models in both the calibration and validation phases, while DC and CC were used to evaluate the simulation results in terms of agreement between the experimental and predicted values. NF and MLP showed a strong agreement between the simulated and experimental tR values. With DC values of 0.9999 and 0.9999 in the calibration and validation phases, respectively, NF exhibited higher performance capabilities than the other three approaches according to a close comparative examination of Table 1.

The ability of these models to handle highly chaotic and complex nonlinear data is demonstrated by the better goodness-of-fit information for DC and CC and the lower RMSE and MSE performance error information for NF and MLP. Performance error of models reporting MSE. A further comparison of performance capabilities revealed that NF improved the prediction

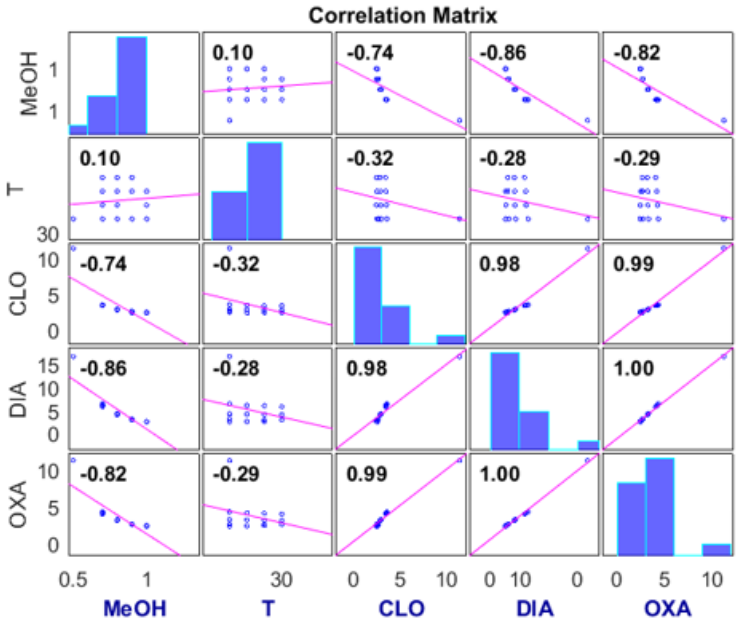

Figure 2. Correlation matrix of the variables.

Table 1. Performance of four different techniques in modelling Clonazepam.

	Calibration				Verification			
	DC	**CC**	**MSE**	**RMSE**	**DC**	**CC**	**MSE**	**RMSE**
NF	0.9999	0.9999	7.60559E-10	2.76E-05	0.9999	0.9999	7.81E-10	2.79E-05
HW	0.647846	0.804889	0.031259399	0.176803	0.6387	0.799187	0.043567	0.208727
MLP	0.999939	0.99997	5.39E-06	0.002321	0.824609	0.90808	9.01E-06	0.003002
SWLR	0.637429	0.798391	0.032184117	0.179399	0.734241	0.856879	1.37E-05	0.003696

accuracy of HW, MLP, and SWLR by 36.03%, 17.53%, and 26.57%, respectively, in the validation phase.

The statistical indices (R2, CC, MSE, and RMSE) used to evaluate the performance prediction of the proposed study were sufficient to assess the effectiveness of the model, taking into account both errors and goodness-of-fit standards. As shown in Table 2, it was possible to observe how well each model performed in terms of the statistical criteria. With an average performance efficiency of 98% compared to 47% for the linear model, the performance results demonstrate the reliability of the non-linear models NF, HW, and MLconcerning SWLR.

To compare the performance results obtained in the current study with those reported in recent literature, Pasin et al. utilised an MLP model to predict the retention time of various new psychoactive substances (NPS), achieving DC values of 0.942. In contrast, the results from our current work using MLP and NF models surpassed those reported by Pasin et al., as indicated by higher DC values (Pasin, Mollerup, Rasmussen, Linnet, & Dalsgaard, 2021). Additionally, previous research implemented a genetic algorithm (GA), multiple linear regression (MLR), and partial least squares (PLS), as well as nonlinear regression techniques, such as kernel PLS (KPLS) and the Levenberg–Marquardt artificial neural network (L–M ANN), to model the retention times of different sedative agents. The correlation coefficients (CC) of the GA-KPLS and L–M ANN models for the training and test sets were 0.921, 0.960, 0.892, and 0.925, respectively. In comparison, the best-performing model in our study, namely NF, demonstrated an exceptional average CC value of 0.9999 in both the training and testing sets for the three psychoactive agents (Noorizadeh, & Noorizadeh, 2012).

Furthermore, Lee et al. applied three models—ANN, support vector machine (SVM), and k-nearest neighbour (k-NN)—to predict the retention times of unknown controlled substances and new psychoactive substances. Using 193 LC–MS-MS barcode spectra as an external test set, the accuracy of the ANN,

Table 2. Performance of four different data-driven technique in modelling Diazepam.

	Calibration				Verification			
	DC	**CC**	**MSE**	**RMSE**	**DC**	**CC**	**MSE**	**RMSE**
NF	0.9998	0.9999	7.968E-10	2.82E-05	0.999997	0.999999	5.41E-10	2.33E-05
HW	0.99911	0.999555	8.47124E-05	0.009204	0.477211	0.690805	0.000108	0.01039
MLP	0.985413	0.99268	0.001389049	0.03727	0.964295	0.981985	7.37E-06	0.002715
SWLR	0.47615	0.690036	0.049883222	0.223346	0.216053	0.464815	0.000162	0.012724

SVM, and k-NN models was 72.5%, 90.0%, and 94.3%, respectively. However, these figures are comparatively lower than those obtained in our current study, particularly for our best-performing models, namely, the NF and MLP (Lee et al., 2022). Hence, based on the literature review, it is evident that the NF and MLP models are robust and serve as effective tools for predicting retention times.

The performance of the models was illustrated using a scatter plot to show the performance errors according to the goodness of fit and the corresponding RMSE values (Figure 3).

The results demonstrate that the three machine learning-based models outperform the traditional SWLR in predicting the tR of the DIA in terms of both goodness of fit and performance error. In addition, the Taylor diagram provides a useful visual tool for comparing the performance of the models (Figure 4). Several statistical measures are described in the Taylor diagram, including standard deviation, CC, DC, and RMSE. Due to its wide range of applications, the Taylor diagram has been used in numerous disciplines, including hydrological modelling, wastewater management, and water engineering. As shown in Figure 4, all machine learning techniques (NF, MLP, and HW) demonstrated greater fitness throughout both the calibration and verification phases. The graph generally indicates that all models demonstrated higher performance and can be used to model the DIA.

Table 3 presents a comparison of four different data-driven methods (MLP, NF, HW, and SWLR models) for OXA simulation in the development of HPLC methods. The results indicate that the MLP model has the most robust predictive capability, marginally surpassing the NF model based on the evaluation parameters employed in this study. The hierarchical order MLP>ANFIS>SWLR>HW can be used to demonstrate the performance effectiveness of the models. In this study, the performance of the models was examined using the CC evaluation metric in both the calibration and validation phases.

Consequently, the principal limitations of the current study include the limited training data and the complexity of retention behaviour. The retention behaviour of psychoactive agents in HPLC is influenced by multiple factors, such as compound structure, polarity, ionisation state, and interactions with the stationary and mobile phases. Accurately capturing the full complexity of retention behaviour in a machine learning model requires advanced feature engineering and sophisticated algorithmic approaches.

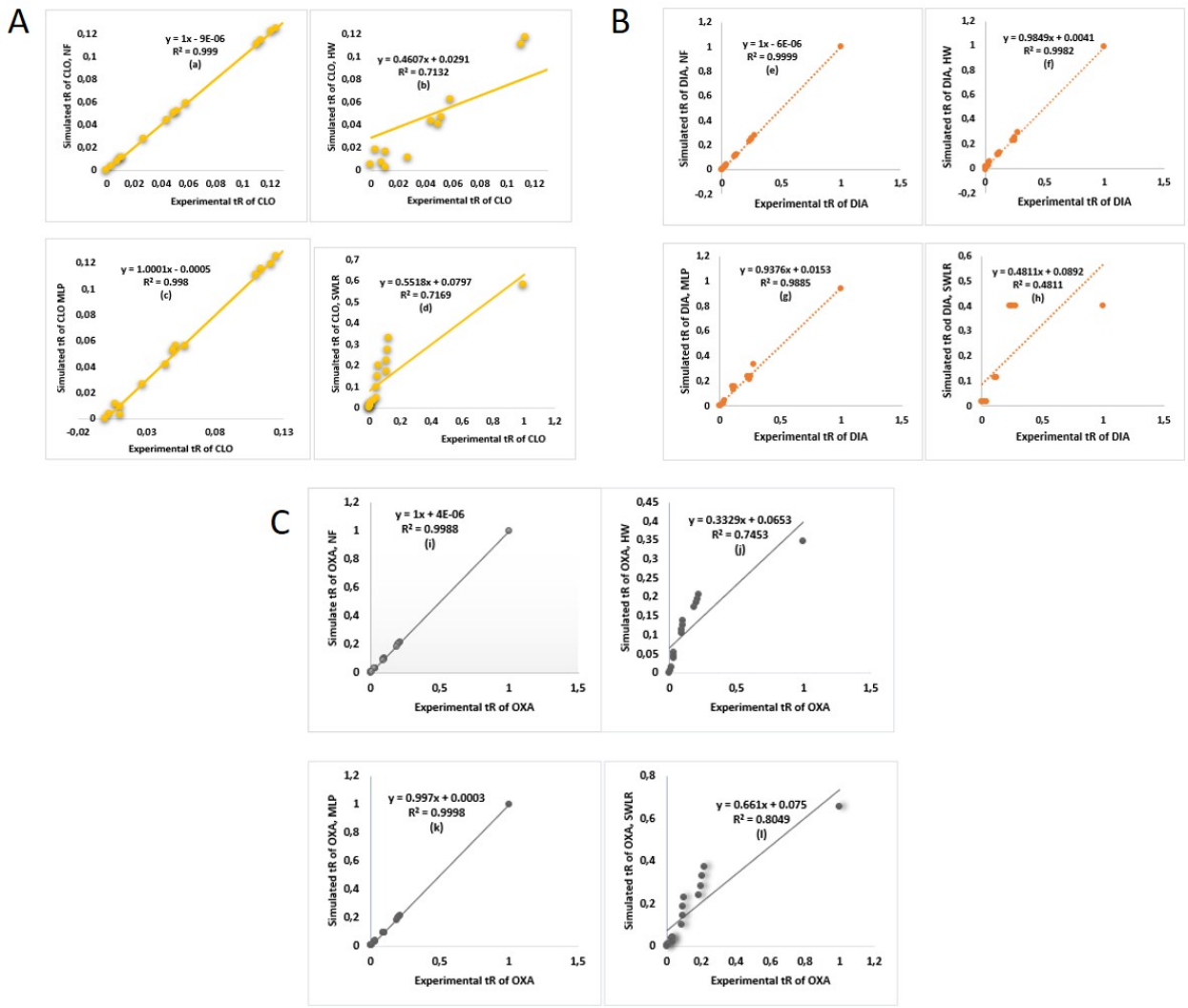

Figure 3. Scatter plots of NF, HW, MLP, and SWLR modelling for tR of CLO (A), DIA (B), and OXA (C), respectively.

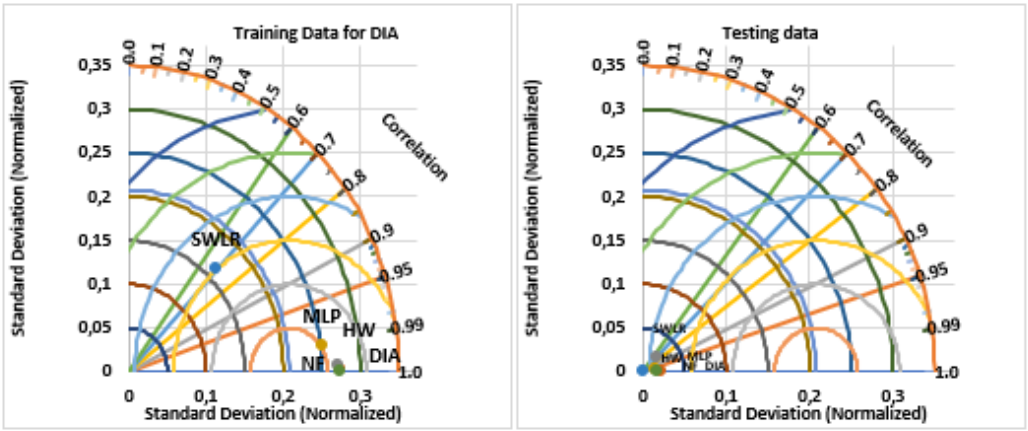

Figure 4. Taylor plot modelling for DIA.

Table 3. Performance of four different data-driven techniques in modelling Oxazepam.

	Calibration				Verification			
	DC	**CC**	**MSE**	**RMSE**	**DC**	**CC**	**MSE**	**RMSE**
NF	0.9988	0.9998	1.84858E-09	4.3E-05	0.999999	0.999999	1.68E-10	1.3E-05
HW	0.470921	0.686237	0.04790947	0.218882	0.221635	0.470781	9.78E-05	0.009891
MLP	0.999881	0.999941	1.07481E-05	0.0032	0.923293	0.960882	9.64E-06	0.003105
SWLR	0.759768	0.871647	0.021753598	0.147491	0.764478	0.874345	2.96E-05	0.005441

CONCLUSION

The effects of different chromatographic settings often influence the estimation of different analytes. In this study, the retention characteristics of three different psychoactive drugs (CLO, DIA, and OXA were modelled under two different mobile phase and column temperature . The performances obtained by the agents demonstrate the accuracy of the machine learning techniques, in particular NF and MLP, which predicted all three drugs with superior performance in both training and testing phases using four different evaluation criteria. The correlation matrix revealed a significant inverse relationship between the dependent variable tR and the mobile phase, leading to the mobile phase as the dominant parameter of column temperature.

In conclusion, the models proposed by MLP and NF can simulate the retention behaviour of each of the three agents. Furthermore, it is recommended to use additional models and optimisation techniques such as support vector machine (SVM), extreme learning machine (ELM), Harris Hawks optimisation (HHO) technique, genetic algorithms (GA), and particle swarm optimisation (PSO), to improve tR in psychoactive drug prediction.

Acknowledgments: The authors appreciate the work by Jouyban et al. (2009) and acknowledge their permission to use the curated data from their study (https://doi.org/10.1002/jssc.200900389) in our current research.

Peer-review: Externally peer-reviewed.

Author Contributions: Conception/Design of Study- A.G.U., E.E., S.I.; Data Acquisition- A.G.U., E.E., S.I.; Data Analysis/Interpretation- A.G.U., E.E., S.I.; Drafting Manuscript- A.G.U., E.E., S.I.; Critical Revision of Manuscript- A.G.U., E.E., S.I.; Final Approval and Accountability- A.G.U., E.E., S.I.

Conflict of Interest: The authors have no conflict of interest to declare.

Financial Disclosure: The authors declared no financial support.

ORCID IDs of the authors

Abdullahi Garba Usman 0000-0001-5660-4581
Emine Erdağ 0000-0002-1431-935X
Selin Işık 0000-0001-7601-3746

REFERENCES

Abba, S.I., Hadi, S.J., & Abdullahi, J. (2017). River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. *Procedia Computer Science, 120*, 75-82. https://doi.org/10.1016/j.procs.2017.11.212

Abba, S. I., Hadi, S. J., Sammen, S. S., Salih, S. Q., Abdulkadir, R. A., Pham, Q. B., Yaseen, Z. M. (2020). Evolutionary computational intelligence algorithm coupled with a self-tuning predictive model for water quality index determination. *Journal of Hydrology, 587*, 124974. https://doi.org/10.1016/j.jhydrol.2020.124974

Abba, S., Usman, A., & I, 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, 201*, 104007. https://doi.org/10.1016/j.chemolab.2020.104007

Abba, S. I., Linh, N. T. T., Abdullahi, J., Ali, S. I. A., Pham, Q. B., Abdulkadir, R. A., ... & Anh, D. T. (2020). Hybrid machine learning ensemble techniques for modeling dissolved oxygen concentration. *IEEE Access, 8*, 157218-157237. https://doi.org/10.1109/ACCESS.2020.3017743

Chandwani, V., Vyas, S. K., Agrawal, V., & Sharma, G. (2015). Soft computing approach for rainfall-runoff modelling: a review. *Aquatic Procedia, 4*, 1054-1061. https://doi.org/10.1016/j.aqpro.2015.02.133

Choubin, B., Khalighi-Sigaroodi, S., Malekian, A., & Kişi, Ö. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal, 61*(6), 1001-1009. https://doi.org/10.1080/02626667.2014.966721

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

D'Archivio, A. A. (2019). Artificial neural network prediction of retention of amino acids in reversed-phase HPLC under application of linear organic modifier gradients and/or pH gradients. *Molecules, 24(3)*, 632. https://doi.org/10.3390/molecules24030632

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*, 123962. https://doi.org/10.1016/j.jhydrol.2019.123962

Erdag, E., Haskologlu, I. C., Mercan, M., Abacioglu, N., & Sehirli, A. O. (2023). An in silico investigation: Can melatonin serve as an adjuvant in NR1D1-linked chronotherapy for amyotrophic lateral sclerosis?. *Chronobiology International, 40*(10), 1395–1403. https://doi.org/10.1080/07420528.2023.2265476

Fine, J., Mann, A. K. P., & Aggarwal, P. (2024). Structure Based Machine Learning Prediction of Retention Times for LC Method Development of Pharmaceuticals. *Pharmaceutical Research, 41*(2), 365–374. https://doi.org/10.1007/s11095-023-03646-2

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), 1-12. https://doi.org/10.1007/s42452-020-03690-2

Ghorbani, M.A., Deo, R.C., Yaseen, Z.M., H. Kashani, M., Mohammadi, B. (2017). Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran.*Theoretical and Applied Climatology, 133* , 1119–1131. https://doi.org/10.1007/s00704-017-2244-0

Hadi, S. J., Abba, S. I., Sammen, S. S., Salih, S. Q., Al-Ansari, N., & Yaseen, Z. M. (2019). Non-linear input variable selection approach integrated with non-tuned data intelligence model for streamflow pattern simulation. *IEEE Access, 7*, 141533-141548. https://doi.org/10.1109/ACCESS.2019.2943515

Işık, S., Usman, A., & 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. İstanbul *Journal of Pharmacy, 53*(3), 320-328. https://doi.org/10.26650/IstanbulJPharm.2023.1181298

Jouyban, A., Soltanpour, S., Acree Jr, 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

Karimi, S., Kisi, O., Shiri, J., & Makarynskyy, O. (2013). Neuro-fuzzy and neural network techniques for forecasting sea level in Darwin Harbor, Australia. *Computers & Geosciences, 52*, 50-59. https://doi.org/10.1016/j.cageo.2012.09.015

Kim, S., & Singh, V. P. (2014). Modeling daily soil temperature using data-driven models and spatial distribution. *Theoretical and Applied Climatology, 118*(3), 465-479. https://doi.org/10.1007/s00704-013-1065-z

Lee, J. K., Han, W. S., Lee, J. S., & Yoon, C. N. (2017). A Novel Computational Method for Biomedical Binary Data Analysis: Development of a Thyroid Disease Index Using a Brute-Force Search with MLR Analysis. *Bulletin of the Korean Chemical Society, 38*(12), 1392-1397. https://doi.org/10.1002/bkcs.11308

Lee, S. Y., Lee, S. T., Suh, S., Ko, B. J., & Oh, H. B. (2022). Revealing unknown controlled substances and new psychoactive substances using high-resolution LC–MS-MS machine learning models and the hybrid similarity search algorithm. *Journal of Analytical Toxicology, 46*(7), 732-742. https://doi.org/10.1093/jat/bkab098

Noorizadeh, H., & Noorizadeh, M. (2012). QSRR-based estimation of the retention time of opiate and sedative drugs by comprehensive two-dimensional gas chromatography. *Medicinal Chemistry Research, 21*, 1997-2005. https://doi.org/10.1007/s00044-011-9727-9

Osipenko, S., Bashkirova, I., Sosnin, S., Kovaleva, O., Fedorov, M., Nikolaev, E., & Kostyukevich, Y. (2020). Machine learning to predict retention time of small molecules in nano-HPLC. *Analytical and Bioanalytical Chemistry, 412*(28), 7767-7776. https://doi.org/10.1007/s00216-020-02905-0

Pasin, D., Mollerup, C. B., Rasmussen, B. S., Linnet, K., & Dalsgaard, P. W. (2021). Development of a single retention time prediction model integrating multiple liquid chromatography systems: application to new psychoactive substances. *Analytica Chimica Acta, 1184*, 339035. https://doi.org/10.1016/j.aca.2021.339035

Pham, Q. B., Abba, S. I., Usman, A. G., Linh, N. T. T., Gupta, V., Malik, A., ... & Tri, D. Q. (2019). Potential of hybrid data-intelligence algorithms for multi-station modelling of rainfall. *Water Resources Management, 33*(15), 5067-5087. https://doi.org/10.1007/s11269-019-02408-3

Rodríguez-Díaz, J. M., & Sánchez-León, G. (2019). Optimal designs for multiresponse models with double covariance structure. *Chemometrics and Intelligent Laboratory Systems, 189*, 1-7. https://doi.org/10.1016/j.chemolab.2019.03.009

Sahu, P. K., Ramisetti, N. R., Cecchi, T., Swain, S., Patro, C. S., & Panda, J. (2018). An overview of experimental designs in HPLC method development and validation. *Journal of Pharmaceutical and Biomedical Analysis, 147*, 590-611. https://doi.org/10.1016/j.jpba.2017.05.006

Samuelsson, J., Eiriksson, F. F., Åsberg, D., Thorsteinsdóttir, M., & Fornstedt, T. (2019). Determining gradient conditions for peptide purification in RPLC with machine-learning-based retention time predictions. *Journal of Chromatography A, 1598*, 92-100. https://doi.org/10.1016/j.chroma.2019.03.043

Shojaeimehr, T., & Rahimpour, F. (2018). Retention time modeling of short-chain aliphatic acids in aqueous ion-exclusion chromatography systems under several conditions using computational intelligence methods (artificial neural network and adaptive neuro-fuzzy inference system). *Journal of Liquid Chromatography & Related Technologies, 41*(12), 810-817. https://doi.org/10.1080/10826076.2018.1518846

Silva, J. M., Azcárate, F. J., Knobel, G., Sosa, J. S., Carrizo, D. B., & Boschetti, C. E. (2020). Multiple response optimization of a QuEChERS extraction and HPLC analysis of diclazuril, nicarbazin and lasalocid in chicken liver. *Food Chemistry, 311*, 126014. https://doi.org/10.1016/j.foodchem.2019.126014

Sultanoglu, N., Erdag, E., & Ozverel, C. S. (2023). A single antiviral for a triple epidemic: is it possible?. *Future Virology, 18*(10), 633-642. https://doi.org/10.2217/fvl-2023-0048

Tayyebi, S., Hajjar, Z., & Soltanali, S. (2019). A novel modified training of radial basis network: Prediction of conversion and selectivity in 1-hexene dimerization process. *Chemometrics and Intelligent Laboratory Systems, 190*, 1-9. https://doi.org/10.1016/j.chemolab.2019.05.005

Tomić, J., Ivković, B., Oljačić, S., Nikolić, K., Maljurić, N., Protić, A., & Agbaba, D. (2020). Chemometrically assisted RP-HPLC method development for efficient separation of ivabradine and its eleven impurities. *Acta Chromatographica, 32*(1), 53-63. https://doi.org/10.1556/1326.2019.00659

Usman, A. G., Işik, S., & Abba, S. I. (2020). A novel multi-model data-driven ensemble technique for the prediction of retention factor in HPLC method development. *Chromatographia, 83*(8), 933-945. https://doi.org/10.1007/s10337-020-03912-0

Usman, A. G., Işik, S., & Abba, S. I. (2021). Hybrid data-intelligence algorithms for the simulation of thymoquinone in HPLC method development. *(7)*Journal of the Iranian Chemical Society, 18, 1537-1549. 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, 5*(10), 579-587. 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, 44*(4), 843-849. https://doi.org/10.1002/jssc.202000890

Zhang, Y., Liu, F., Li, X. Q., Gao, Y., Li, K. C., & Zhang, Q. H. (2024). Generic and accurate prediction of retention times in liquid chromatography by post–projection calibration. *Communications Chemistry, 7*(1), 54. https://doi.org/10.1038/s42004-024-01135-0

How cite this article

Usman, A.G., & Erdağ, E., & Işık, S. (2024). Effects of chromatographic conditions on retention behaviour of different psychoactive agents in high-performance liquid chromatography: A machine-learning-based approach. *İstanbul Journal of Pharmacy, 54*(2), 133–143. DOI: 10.26650/IstanbulJPharm.2024.1225463