Development Of Computational Intelligence Algorithms For Modelling The Performance Of Humanin And Its Derivatives In HPLC Optimization Method Development

Umar Muhammad Ghali, Mohamed Alhosen Ali Degm, Ahmed Nouri Alsharksi, Qendresa Hoti, Abdullahi Garba Usman

Abstract: Humanin and its derivatives are considered as neural cells protecting agents against pathological proteins such as the amyloid protein precursor that causes the alzheimer's disorder. The precise prediction of the properties of humanin in high performance liquid chromatography (HPLC) optimization method is of paramount importance. Therefore, to achieve this the development of resilient and satisfactory computational intelligence tools is crucial. In the current study, the comparative potential performance of adaptive neurofuzzy inference system (ANFIS) and multilinear regression models. The outputs given by the ANFIS and MLR models were compared with the experimental values through two statistical evaluation indices Nash-Sutcliffe efficiency (NC) and Mean squared error (MSE). Graphical illustrations such as scatter plot and time series were employed to compare the performance of the models. The results of the study indicated that ANFIS outperformed MLR for predicting the maximum retention time (tR max) and resolution of humanin and its derivatives in HPLC optimization method development. Equally, ANFIS showed the highest value of NC (0.9999/ 0.9992) for tR max and (0.9998/ 0.9994) for resolution in the training and testing stages respectively. Similarly, ANFIS indicated lowest values of MSE for tR max and resolution in both the training and testing stages. The comparative analysis of the result demonstrated that ANFIS as a promising non-linear artificial intelligence based model found to be more reliable and suitable for predicting the performance of humanin and its derivatives in HPLC optimization method development.

Keyword: Humanin; HPLC; ANFIS; MLR; retention time; resolution

1. INTRODUCTION

Generally, humanin as well as its derivatives are considered as neural cells protecting agents against pathological proteins such as the amyloid protein precursor that causes the Alzheimer's disorder. Alzheimer's disease is one of the neurodegenerative diseases that composed of cognitive disorders as well as dementia. The curing of this disease is of vital using agents such as humanin and its derivatives [1]. Therefore, the analytical determination of humanin and its derivatives is of paramount importance. Even though these agents have been studies and examined in the literature. Regarding the peptide bond separation, the reverse phase ion-pair high-pressure liquid chromatography (HPLC) with the used of trifluoroacetic acid (TFA) as the mobile phase is usually used to serve as an alternative over the ion-exchange liquid chromatography [2]. In the optimization method using HPLC technique, the vital and significant aspect is the achievement of sufficient separation for all the derivatives within a sound time. Longer analysis time is the major problem of elution, which can be resolved using a gradient elution system. Generally, many samples cannot be separated successfully through the applications of isocratic conditions but rather through gradient elution method (also known as solvent programming): which involves changing the composition of mobile phase during the separation process in order to progressively reduce the retention of the sample [3]. Simulation with regard to the optimization of various chromatographic conditions such as the composition of the mobile phase (concentration of organic modifier), pH, and the temperature is important to achieve a good resolution [4]. The mathematical approach to accomplish such a goal is defined as optimization. This method involves the optimization of pH, which gives better selectivity due to the degree of ionization of the solutes, the mobile phase, and stationary phase additives, which might affect the pH. It also focuses on optimizing the composition of the mobile phase, which is the ratio of the water and the organic modifier. Moreover, the optimization various chromatographic factors such as the composition of the mobile phase and column temperature are very important in obtaining sufficient resolution. Considering the high number of factors that influences separation, it is very difficult as well as time consuming in reaching the desired separation conditions, more especially employing the single parameter optimization method, which involves changing of one parameter in a time while other variables are kept constant. Now a day, the global optimization approach through coupling the experimental and data driven approaches has been used. This coupling is done according to the suitable experimental design, whereby the optimal separation conditions are simulated using artificial intelligence (AI) based models by employing the output and input variables of the experimental studies. The applications of AI based models in the area of chromatography has been reported in the literature. For example, Zeng et al., described the qualitative prediction of a rare Chinese herbal medicine known as Tetrastigma Hemsleyanum using a combination of principal component analysis (PCA) and Adaboost method with logistic regression (Adaboost -LR) in HPLC method development. The combination proves the reliability of these models in determining the origin of this herbal plant [5]. Also S. Agatonovic-Kustrin et al., [6] reported the use of neural networks for HPLC optimization response surface models compared with multiple regression (MLR) methods. Many studies based on the applications of data driven models on HPLC optimization method development can equally be in [7],[8],[9], [10], [11], [12], [13], [14], [15], [16], [17] from the literature. Considering the previous related studies conducted in the technical literature, it is a proof that AI-based models are promising, satisfactory and reliable tools in chemical and chromatographic modelling methods. This study employs the comparative application of

AI-based model inform of adaptive neuro-fuzzy inference system and a classical linear model (MLR) for modelling the maximum retention time (tRmax) and resolution of humanin and its derivatives in HPLC optimization method development . The study involves the use of mobile phase inform of acetonitrile (ACN) and trifluoroacetic acid (TFA) and column temperature as the input parameters.

2. MATERIAL AND METHODS

2.1 Proposed methodology

In this study, an AI-based model informs of adaptive neuro-fuzzy inference system and a classical linear model (MLR) for modelling the maximum retention time (tRmax) and resolution of humanin and its derivatives in HPLC optimization method development. The data of this study was collected from previous experimental studies conducted by [1]. In this context, tRmax and resolution of humanin derivatives were modelled by employing various input variables inform of concentrations of acetonitrile (ACN) and trifluoroacetic acid (TFA) and column temperature. Though other chromatographic variables can be employed or be simulated using similar approach. Practically, it is hard to determine a single data driven approach that has higher performance over other models in simulating various chromatographic properties. It is therefore very difficult to choose a specific model in a certain study by modelers. The motivation of this proposed technique is to predict the maximum retention time (tRmax) and resolution of humanin and its derivatives in HPLC optimization method development using the concentrations of acetonitrile (ACN) and tsrifluoroacetic acid (TFA) and column temperature as the corresponding input variables. Fig. 1 demonstrated the flowchart of the proposed methodology.

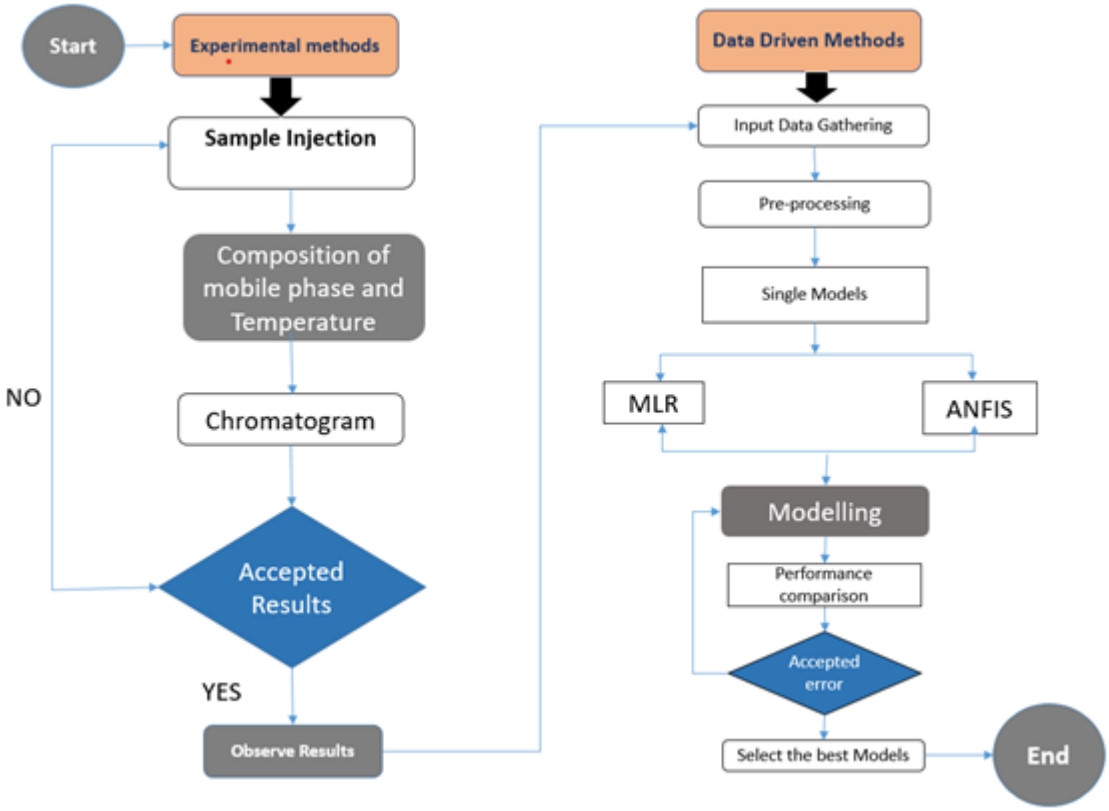


***Fig. 1**. Flowchart of the proposed methodology*

2.2 Adaptive neuro-fuzzy inference system (ANFIS)

Even though, ANNs tools are one of the broadly use AI-based model which is motivated by copying the brain of human being, as a result of its resilience of mimicking with a high complex connection between the input and output models of the data collections [18]. ANFIS has been demonstrated to be a successful software that incorporates the approach of the fuzzy Sugeno model that benefits from both fuzzy logic and ANN in one system. ANFIS has been recently used in predicting and modelling complex datasets [7].ANFIS is also a real-world estimator because of its capacity to approximate real functions. Fuzzy logic converts the input data into fuzzy values via the application of membership functions. The numbers range between 0-1 [19]. Furthermore, in the ANFIS model, nodes work as membership functions (MFs) and also allow the modelling between the relations of the input with the output. Assume the FIS contains two inputs 'x' and 'y' and one output 'f', a first-order Sugeno fuzzy has the following rules.

Rule 1: if μx is A1 and μy B1then f1= p1x+q1y+r1 (2)

Rule 2: if μx is A2 and μy is B2 then f2= p2x+q2y+r2 (3)

A_1, B_1,A_2,B_2 Parameters are membership functions for x and y inputs

p_1,q_1,r_(1,) p_2,q_2,r_(2,) are outlet function parameters. The structure and formulation of ANFIS follows a five-layer neural network arrangement.

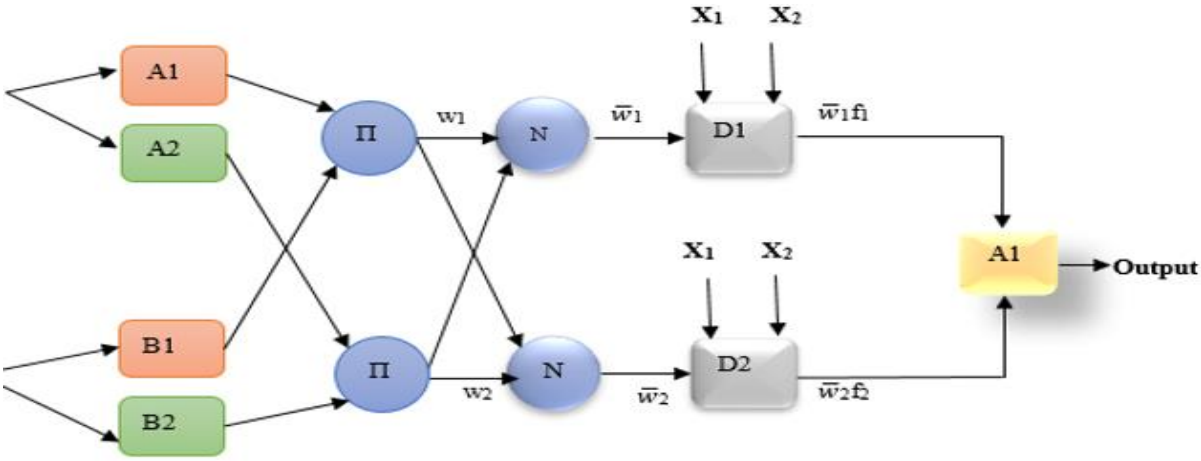


***Fig. 2**. The architecture illustration of ANFIS model*

2.3 Multi-linear regression (MLR)

Multi-linear regression (MLR) is the commonest linear method employed by modelers in predicting different variables as used in different field of study in science, engineering, health science and social sciences. It aids in understanding the linear relationship between the predictor and the input variables [20]. It explores the interaction between the variables and describes the relationship between them by keeping the independent variables fixed and varying one. Nourani et al., (2020a), can correlate the n regressor variables and the dependent variable y:

y= b_0+ b_1 x_1+ b_2 x_2+ b_3 x_3+⋯+ b_i x_i+ξ

(4)

In Equation 4, x_i represents value of the i^th predictor, b_istands for coefficient of the i^th predictor, b_0 is the constant of regression and ξ is the error term.

2.4 Performance criteria indices and validation method for data-driven models

Usually, for any form of data-driven approach, the performances of the models are evaluated through different criteria based on a comparison between the simulated and experimental values. In this work, the Nash-Sutcliffe coefficient (NC) as a goodness-of-fit (RMSE) and mean-squared error (MSE), were used for the evaluation of the models:

$$NC = 1 - \frac{\sum_{j=1}^{N}[(Y)_{obs,j} - (Y)_{com,j}]^2}{\sum_{j=1}^{N}[(Y)_{obs,j} - \overline{(Y)}_{obs,j}]^2}$$

(5)

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_{obsi} - Y_{comi})^2$$

(6)

Where N, Y_{obsi}, $\bar{Y}$ and Y_{comi} are the data number, observed data, average value of the observed data and computed values, respectively.

For the validation technique, different types of validation methods can be applied such as cross-validation (i.e., k-fold cross validation), holdout and leave one out. In this work the k-fold cross-validation is used, which is regarded as the process employed in order to reduce the problems of overfitting. In this technique, the initial data set is categorized into same-sized subsets of k [21]. From the k-1 data subsets, one will be retained and employed for validation purposes and the remaining subsets will be maintained and utilized for training purposes [22]. Thereby, the validation will be repeated in k-folds [20]. The result of these alternations is considered as the average of the validation efficiency of the k-subsets. Generally, k-values are determined via the sample availability, mostly from 2-10. One of the major advantages of the k-fold cross validation process is that in each round, the training and the validation set are independent from each another. This leads to a performance objective that provides a good foundation for the model optimization [23].

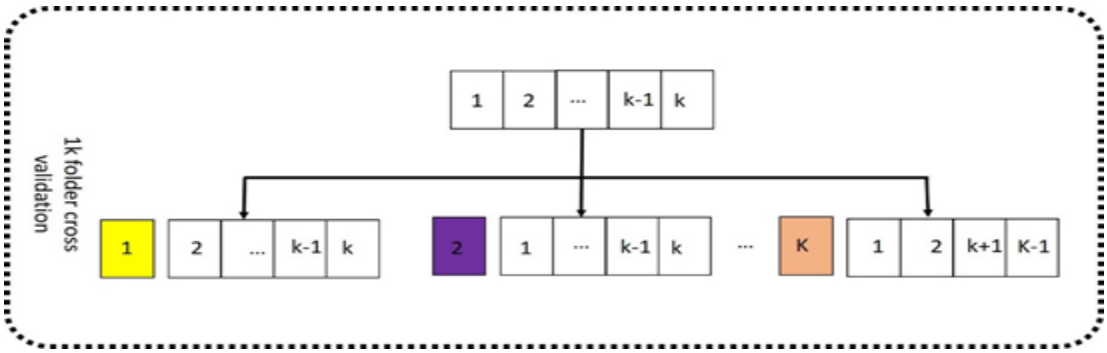


***Fig. 3**. Demonstration showing the k-fold cross-validation [24].*

3. APPLICATION OF RESULTS AND DISCUSSION

The models (MLR and ANFIS) were analysed in order to predict the qualitative properties of humanin and its derivatives in HPLC optimization method. Before the training of the data in the models, the data set were statistically evaluated as seen in Table 1, these analysis were done generally in order to understand the nature and behaviour of the data prior to the modelling.

Table 1. Correlation and statistical analysis

Correlation Analysis

	ACN	TFA	Tem.	tR (max)	Resolution
ACN	1				
TFA	0.001912	1			
Tem.	0	0	1		
tR (max)	-0.58953	0.667744	-0.10871	1	
Resolution	-0.49755	0.596748	0.114871	0.469788	1

Statistical Analysis

	ACN	TFA	T	tR (max)	Resolution
Mean	32	0.0697	35	18.2131	14.5043
Median	32	0.07	35	11.375	15.46
Standard Deviation	1.9169	0.0187	11.4076	17.4756	3.9957
Minimum	28.6	0.036	15	2.98	6.48
Maximum	35.4	0.1	55	62.2	20.96

The correlation analysis was employed in order to demonstrate the relation that exists between the parameters, which is shown using a linear function. The correlation strength does not depend on the sign or the direction. Positive sign shows that an increase in the first variable will lead to a direct increase to the second variable, while a negative sign demonstrated an inverse relationship between the first and the second variable. It can be seen from table 1 that there is higher correlation between TFA and the output variable tR (max) and Resolution with a correlation value of R=0.667744 and 0.596748.

3.1 Results of the data driven models

The performance skills of the data driven models are demonstrated in table 1. Based on the comparative analysis of the models, it can be observed that all the two data driven approaches have the ability of predicting the tR (max) and resolution of humanin and its derivatives using HPLC technique. However, ANFIS gives superior and higher alternative as compared with MLR in both the training and testing stages as far as the performance evaluation indices used in the studies NC and MSE are concern. However, the performance accuracy of the models in terms of NC indicated that ANFIS outperformed MLR model and boost its performance accuracy by 26% and 5% for modelling of tR (max) and Resolution respectively in the testing stage.

Table 2: *Results of the MLR and ANFIS models*

	Training		Testing	
	NC	MSE	NC	MSE
ANFIS-tR(max)	0.9999	0.0029	0.9992	0.0991
MLR-tR(max)	0.8484	96.8501	0.7375	31.4717
ANFIS-Reso.	0.9998	0.0337	0.9994	0.0337
MLR-Reso.	0.9732	5.7025	0.9533	5.7025

3.2 Comparison for MLR and ANFIS models

The comparative performance of the results can be equally demonstrated using a scatter plot. From Fig. 4 it can be observe that ANFIS is more robust as compared to MLR in modelling the qualitative performance of humanin using HPLC technique. Close comparative analysis of the plot shows the strong agreement between the experimental and simulated values were associated with the ANFIS model. Further comparison of the mean square error (MSE) is equally shown in Fig. 5. It is worth to mention that the errors from both models (MLR and ANFIS) were within the acceptable range in both the training and testing stages, except in modelling of tR (max) using MLR model, even though the performance of ANFIS is much higher as compared with that of MLR. This verified that the non-linear model has the ability of capturing the complex and chaotic system in the simulation of tR (max) and Resolution in HPLC optimization method development. This was in line with the studies of [25],[26].

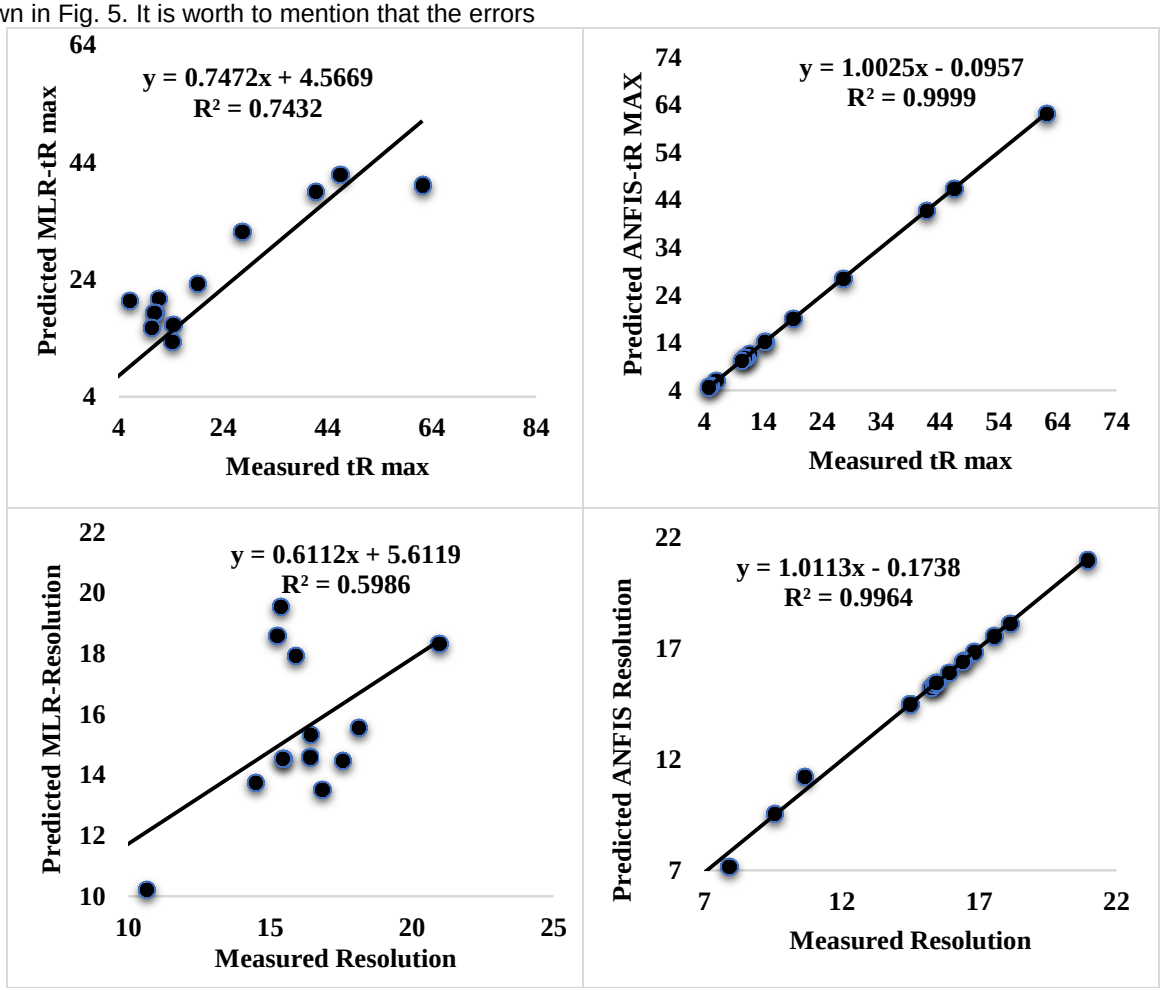


***Fig. 4**. Scatter plots for MLR and ANFIS for tR (max) and Resolution in HPLC optimization method development*

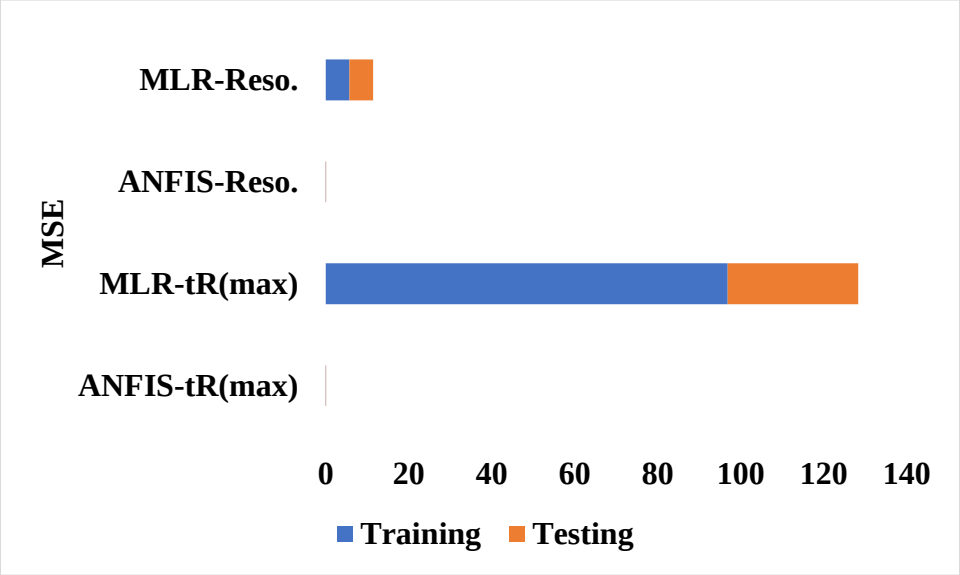
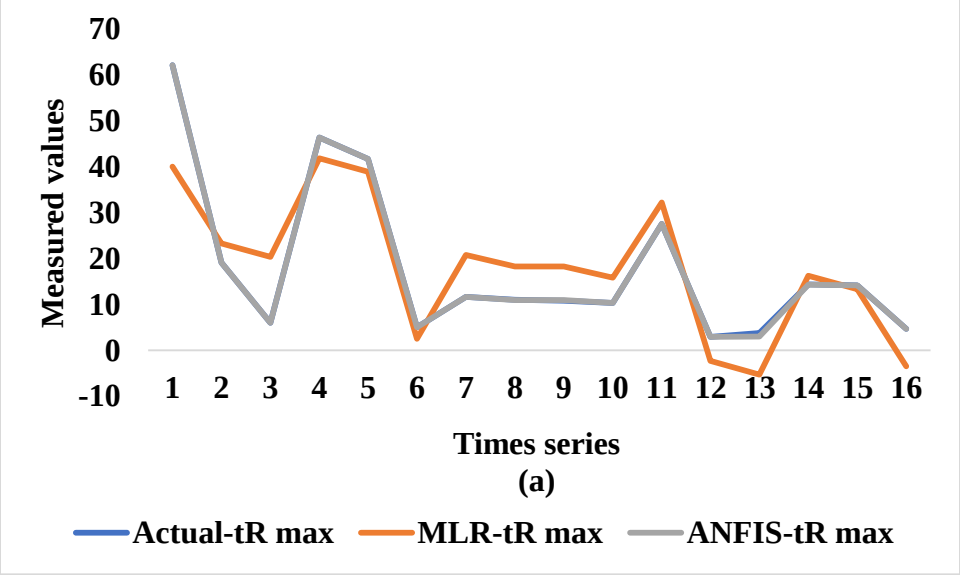


***Fig.5**: Comparative analysis of the performance error of the data driven approaches*

More also, Figure 6 demonstrated the time series response plot for modelling the performance of both tR (max) and resolution in HPLC optimization method development. The extent by which the values were spread i.e between the observed and predicted values has proved table 2 above.



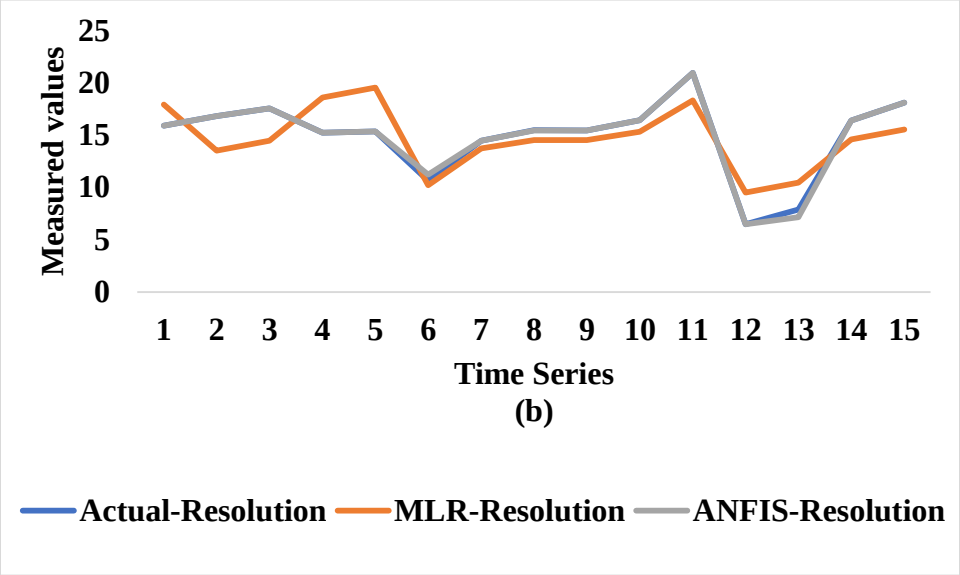


Fig. 6 *Time series for (a) tR (max) and (b) Resolution in HPLC optimization method development*

CONCLUSION

The current work explored two different data driven approaches, consisting of classical linear model (Multilinear regression analysis (MLR)) and a non-linear model inform of Adaptive neurofuzzy inference system (ANFIS) for the simulation of the qualitative properties of humanin and its derivatives using tR (max) and resolution in HPLC optimization method development. The study employed the use of mobile phase and column temperature as the corresponding input parameters. In this work, the data were obtained from previous experimental designed published in the technical literature as discussed in chapter two. Based on the comparative predictive analysis of the models, the result obtained proved the efficiency of ANFIS as a non-linear model over MLR with a considerable and satisfactory performance accuracy. With regard to the goodness of fits in terms of NC the ANFIS model outperformed the MLR model and increase its performance skills by 26% and 5% for tR (max) and resolution respectively. Generally, the artificial intelligence based models displayed higher performance accuracy as compared with the classical linear approaches and are therefore regarded satisfactory and reliable tools for modelling the performance of humanin and its derivatives using HPLC technique. The results equally indicated that other non-linear model, optimization methods as well as other techniques such as principle component analysis (PCA), genetic algorithms (GA), ensemble technique and hybrid data driven algorithms etc can be employed in order to ascertain the performance accuracy of the models. Conflict of interest: The authors declared that there is no conflict of interest in submitting this work

REFERENCES

[1] K. Novotná, J. Havliš, and J. Havel, “Optimisation of high performance liquid chromatography separation of neuroprotective peptides: Fractional experimental designs combined with artificial neural networks,” J. Chromatogr. A, vol. 1096, no. 1–2, pp. 50–57, 2005.

[2] C. Veenaas, A. Linusson, and P. Haglund, “Retention-time prediction in comprehensive two-dimensional gas chromatography to aid identification of unknown contaminants,” Anal. Bioanal. Chem., vol. 410, no. 30, pp. 7931–7941, 2018.

[3] L. R. S. and J. W. Dolan, “High-Performance Gradient Elution: The Practical Application of the Linear-Solvent-Strength Model.” John Wiley & Sons.

[4] S. Agatonovic-Kustrin, M. Zecevic, and L. Zivanovic, “Use of ANN modelling in structure-retention relationships of diuretics in RP-HPLC,” J. Pharm. Biomed. Anal., vol. 21, no. 1, pp. 95–103, 1999.

[5] J. Zeng, Q. Chai, X. Peng, and S. Li, “Geographical Origin Identification for Tetrastigma Hemsleyanum Based on High Performance Liquid Chromatographic Fingerprint,” Proc. - 2019 Chinese Autom. Congr. CAC 2019, pp. 1816–1820, 2019.

[6] S. Agatonovic-Kustrin, M. Zecevic, L. j. Zivanovic, and I. G. Tucker, “Application of artificial neural networks in HPLC method development,” J. Pharm. Biomed. Anal., vol. 17, no. 1, pp. 69–76, 1998.

[7] S. I. Abba, A. G. Usman, and S. Işik, “Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach,” Chemom. Intell. Lab. Syst., p. 104007, 2020.

[8] H. I. El-shorbagy, F. Elsebaei, S. F. Hammad, and A. M. El-brashy, “Optimization and modeling of a green dual detected RP-HPLC method by UV and fluorescence detectors using two level full factorial design for simultaneous determination of sofosbuvir and ledipasvir : Application to average content and uniformity of dosage ,” Microchem. J., vol. 147, no. February, pp. 374–392, 2019.

[9] M. A. Aslam et al., “Nano Biomed Eng SVM Based Classification and Prediction System for Gastric Cancer Using Dominant Features of Saliva,” vol. 12, no. 1, pp. 1–13, 2020.

[10] M. M. Aboulwafa et al., “Journal of Pharmaceutical and Biomedical Analysis Authentication and discrimination of green tea samples using UV – vis , FTIR and HPLC techniques coupled with chemometrics analysis,” J. Pharm. Biomed. Anal., vol. 164, pp. 653–658, 2019.

[11] M. Celeste, E. Galvão, B. Rosa, E. Ferreira, R. Silva, and C. Caldas, “Screening of Mangifera indica L . functional content using PCA and neural networks (ANN),” Food Chem., vol. 273, no. December 2017, pp. 115–123, 2019.

[12] J. Tomi et al., “Chemometrically Assisted RP-HPLC Method Development for Efficient Separation of Ivabradine and its Eleven Impurities,” vol. 32, no. June 2019, pp. 53–63, 2020.

[13] E. Science, “Evaluation of quality of Salvia miltiorrhiza Bunge from different provenances by HPLC-DAD fingerprint combined with Chemometrics Method Evaluation of quality of Salvia miltiorrhiza Bunge from different provenances by HPLC-DAD fingerprint combined with Che,” 2019.

[14] M. Reclo, E. Yilmaz, Y. Bazel, and M. Soylak, “Switchable solvent based liquid phase microextraction of palladium coupled with determination by flame atomic absorption spectrometry,” Int. J. Environ. Anal. Chem., vol. 97, no. 14–15, pp. 1315–1327, 2017.

[15] M. A. Korany, H. Mahgoub, O. T. Fahmy, and H. M. Maher, “Application of artificial neural networks for response surface modelling in HPLC method development,” J. Adv. Res., vol. 3, no. 1, pp. 53–63, 2012.

[16] J. Yang, G. Xu, H. Kong, Y. Zheng, T. Pang, and Q. Yang, “Artificial neural network classification based on high-performance liquid chromatography of urinary and serum nucleosides for the clinical diagnosis of cancer,” J. Chromatogr. B Anal. Technol. Biomed. Life Sci., vol. 780, no. 1, pp. 27–33, 2002.

[17] V. Nourani, H. Hakimzadeh, and A. B. Amini, “Implementation of artificial neural network technique in the simulation of dam breach hydrograph,” J. Hydroinformatics, vol. 14, no. 2, p. 478, 2012.

[18] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” Int. J. Forecast., vol. 14, no. 1, pp. 35–62, 1998.

[19] S. I. Abba et al., “Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant,” J. Water Process Eng., vol. 33, no. October 2019, p. 101081, 2020.

[20] F. Khademi and K. Behfarnia, “Evaluation of Concrete Compressive Strength Using Artificial Neural Network and Multiple Linear Regression Models,” Iust, vol. 6, no. 3, pp. 423–432, 2016.

[21] Q. B. Pham et al., “Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall,” Water Resour. Manag., vol. 33, no. 15, 2019.

[22] K. Zarei, M. Atabati, and M. Ahmadi, “Shuffling cross–validation–bee algorithm as a new descriptor selection method for retention studies of pesticides in biopartitioning micellar chromatography,” J. Environ. Sci. Heal. - Part B Pestic. Food Contam. Agric. Wastes, vol. 52, no. 5, pp. 346–352, 2017.

[23] P. Kazemi et al., “Computational intelligence modeling of granule size distribution for oscillating milling,” Powder Technol., vol. 301, pp. 1252–1258, 2016.

[24] H. Sanikhani and O. Kisi, “River Flow Estimation and Forecasting by Using Two Different Adaptive Neuro-Fuzzy Approaches,” Water Resour. Manag., vol. 26, no. 6, pp. 1715–1729, 2012.

[25] E. Dehghanian, M. Kaykhaii, and M. Mehrpur, “Comparison of single best artificial neural network and neural network ensemble in modeling of palladium microextraction,” Monatshefte fur Chemie, vol. 146, no. 8, pp. 1217–1227, 2015.

[26] S. H. Park et al., “Retention prediction of low molecular weight anions in ion chromatography based on quantitative structure-retention relationships applied to the linear solvent strength model,” J. Chromatogr. A, vol. 1486, pp. 68–75, 2017.