ISSN (Print): 2476-8316
ISSN (Online): 2635-3490
Dutse Journal of Pure and Applied Sciences (DUJOPAS), *Vol. 6 No. 2 June 2020*

Modelling the Absorbance of a Bioactive Compound in HPLC Method using Artificial Neural Network and Multilinear Regression Methods

Hadiza Usman Abdullahi[1], A.G Usman[2], S.I.Abba[3]*

[1]Department of Chemistry,
Faculty of Science,
Federal University Dutse,
Jigawa - Nigeria

[2]Department of Analytical Chemistry,
Faculty of Pharmacy,
Near East University,
99138 Nicosia, Turkish Republic of Northern Cyprus

[3]Department of Physical Planning Development,
Yusuf Maitama Sule University
Kano, Nigeria

Email: saniisaabba86@gmail.com

Abstract

In this paper, the implementation of two different models were compared including a non-linear Artificial Neural Network (ANN) and a classical Multilinear Regression (MLR) models for the simulation of absorbance. Different range of concentrations of the standard Isoquercetin (ppm) were considered as the input variables while the absorbance data measured from the experimental analysis using High Performance Liquid Chromatography (HPLC) instrument served as the target variable. The simulated result was further evaluated base on two different performance efficiency criteria (determination coefficient (R^2) and root mean square error (RMSE)). The obtained results showed the ability of the models with the minimum R^2-values of 0.9538 from MLR and maximum R^2-values of 0.9993 from ANN in the testing phase, even though ANN model slightly outperformed MLR model for the simulation of all the three different absorbances. The predictive results also demonstrated that both the two models are capable of modelling the absorbance based on the performance efficiency of the model.

Keywords: Isoquercetin, Absorbance, HPLC, Artificial neural network, Multilinear Regression

**Author for Correspondence*

INTRODUCTION

Flavonoids are a large and complex group of polyphenolic compounds widely distributed throughout the plant kingdom e.g. Isoquercetin. Isoquercetin is a common dietary component of fruits, vegetables and beverages, and are usually present in plant tissue as the form of glycosides (Chang et al., 2005). Flavonoids exhibit a multitude of biological activities, such as anti-bacteria, anti-inflammation, anti-allergy and anti-oxidation. Epidemiological studies have shown that the dietary intake of flavonoids is inversely associated with the incidence of coronary heart disease and cancer. Due to their abundance in dietary products and their potential beneficial pharmacological and nutritional effects, the flavonoids are of considerable interest for drug as well as health food supplement (Romani et al., 2002).

Among the useful aspects of developing, a method in liquid chromatography is to achieve a sufficient and reasonable result in a reasonable and optimum analysis time. Longer analysis time is one of the general elution problems which can be resolved mostly using gradient elution system or simply through adjusting some chromatographic conditions so as to achieve the desired response (Agatonovic-Kustrin et al., 1998). The calibration method of this technique involves the use of variable concentrations (0.5, 1, 10, 25, 50, 100, 250, 500, 750, 1000ppm) in order to determine their respective absorbance (peak area).It can be observed from the reported literature, most of the research established using data-driven algorithms employed the application linear statistical approaches (e.g., MLR).

Artificial intelligence methods have shown rapid and quick increasing applications in various fields of prediction in both science, engineering, and management sciences due it is promising ability, accuracy, and fast learning speed (You et al., 20019). For instances, Yang et al., 2002 studied that the applications of neural network for HPLC optimization response surface model compared with multiple regression (MLR) methods. The results showed that the artificial neural network (ANN) offers great possibilities for method development in HPLC. And the predicted retention capacities of the analytes are better than the ones obtained using multiple regression method. A precise description of the retention capacities related to the triazines as well as the mobile phase is achieved using an ANN. Comparable, the result obtained by the ANN is better than the one obtained by QSRR (Ruggieri et al., 2005). Similarly, Korany et al., 2012 reported that the applications of neural network for HPLC optimization response surface model compared with multiple regression methods. The study involves the combined effect of mobile phase composition and pH on a reversed-phase HPLC to determine the retention factors of different analytes. The results were compared and the ANN result was found to be more accurate than multiple regression analysis. The main aim of this paper is to employ and compare the applications of artificial neural network (ANN) and multi linear regression (MLR) for assessing and determining the absorbance of Isoquercetin using different concentrations. The novelty of this study is presented in two forms; first, since the introduction of this data driven approaches this is the first research conducted in the literature to determine the absorbance of any bioactive ingredient using HPLC method development. Second, to the best

knowledge of the authors, this is the first research conducted on the modelling of Isoquercetin using artificial intelligence based models.

MATERIALS AND METHODS

Reagents and chemicals

Methanol (HPLC grade), formic acid and Isoquercetin standard were all purchase from sigma-Aldrich. All reagents were of analytical grade and used without further purification. Distilled and de-ionized water was used for the preparation of all solutions.

Instrumentation

Chromatographic separations were carried out by an Agilent 1200 series HPLC system (USA) equipped with quaternary pump, vacuum degasser, autosampler, column oven, DAD detector and Agilent Chemstation for LC 3D Systems (Rev. B.03.01) software. The column used in the separation was a reversed-phase (Agilent Eclipse XDB-C18. 4.6 mm ID x 150 mm, 5 µm).

Preparation of standard

Working standard solutions for Isoquercetin were freshly prepared by diluting the stock solution in deionized water to give the following concentration 0.5ppm, 1ppm, 10ppm, 25ppm, 50ppm, 100ppm, 250ppm, 500ppm, 750ppm and 1000ppm respectively.

Artificial Neural Network (ANN)

Artificial neural networks (ANN) belong to the learning machine family, which uses a computational approach to develop predictive models for desired parameters by simulating the way biological neurons work in the human brain (Gaya, M.S et al., 2017). Numerous types of ANN have been adopted in the literature. A feedforward multilayer perceptron (FFMLP) was adopted in this study (Abba, S.I. and Elkiran, G., 2017). With this type of ANN, the learning is supervised, which indicates that for a given set of input vectors, the output vectors are provided to the network and the system is expected to adjust its weights using forward and backward calculations to minimize the prediction errors, which is also known as the learning phase (Elkiran, G., Nourani, V., Abba, S.I. and Abdullahi, J., 2018).

The structure of MLP is divided into three layers, which are strongly interconnected with artificial neurons. The initial layer is the input layer x_i where the input signals are stored for a given set of input parameters x_{ai}. Input parameters in this study were the sample concentrations used for the calibration (Nourani, V., Elkiran, G., Abdullahi, J. and Tahsin, A., 2019).

$$x_i = (x_{1i}, x_{2i}, \dots\dots\dots x_{ai}) \qquad i = 1, \dots\dots, n \qquad (1)$$

The final layer is the output layer y_i, where the targeted absorbance, was expected to be predicted in this study.

$$y_i = (y_{1i}, y_{2i}, \dots\dots\dots y_{bi}) \qquad i = 1, \dots\dots, n \qquad (2)$$

The intermediate layer is called the hidden layer and it is devoted to the calculations that formally connect the input layer x_i with the output layer y_i.
A weighted sum of the values of the input variables is computed through the weights that are associated with each connection by eqn. 3

$$y = \sum_{i=1}^{a} w_i x_i + w_0 \quad (3)$$

Where w_i is the weight associated with the ith input parameter, x_i is the data corresponding to the input parameter and w_o is the bias.
The output value is calculated in a forward pass using a transfer function. Numerous transfer functions such as the Heaviside step function, sigmoidal, or hyperbolic tangent have previously been used in the literature (Abba, S.I., Usman, A.G. and Selin, I.Ş.I.K., 2020). Utilization of the activation function is highly dependent on the nature of the dataset and the type of model desired to be developed. In this study, a hyperbolic tangent as expressed in eqn. 4 was adopted.

$$f(a) = \frac{e^{a}-e^{-a}}{e^{a}+e^{-a}} = \frac{2}{1+e^{-2a}} - 1 \quad (4)$$

The abovementioned procedure is an iterative process. The network performs a series of forward and backward calculations to adjust its weights in order to achieve the most accurate predictions to the target values, which is also called the training of the neural network (Abba et al., 2020).

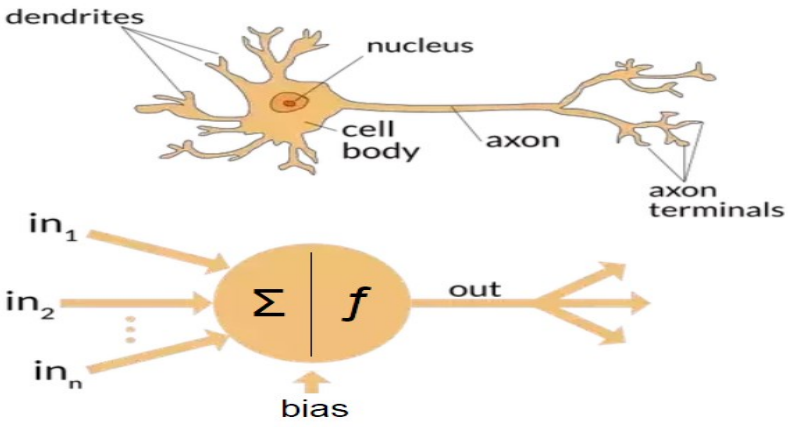


Fig. 1: The three-layer feed forward neural network(Gaya et al., 2017)

Multi-Linear Regression (MLR)

Generally, linear regression is classified into two mains categorized of multiple and simple linear regression. This depends on the aim of the prediction, for example if the aim involves estimating the linear correlation that exists between a single predictor and a single criterion variable. Therefore, such model is considered to be called simple linear regression (SLR). Moreover, if the aim of predicting the linear correlation which exists between more than one predictor and still with a single criterion variable, such model is referred to as Multiple Linear Regression (MLR) (Khademi and Behfarnia, 2016). Mostly, multilinear regression (MLR) is the widely used type of linear regression involves an analysis whereby each value from the independent variables is related to a dependent variable value (Chen and Liu, 2015). Usually, MLR involves the estimation of the level of correlation that exists between a single response variable i.e. the dependent and two or more predictors i.e. independent variables. The overall expression of MLR are presentedin equation 5.

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots b_ix_i \quad (5)$$

Wherex_1, is the value of the i^{th} predictor, b_0 is the regression constant, and $b_i$is the coefficient of the i^{th}predictor.

Furthermore, the models were evaluated using two different performance indices, which include the determination co-efficient (R^2) for goodness of fit and root mean square error (RMSE).

$$R^2 = 1 - \frac{\sum_{j=1}^{N}\left[(Y)_{obs,j} - (Y)_{com,j}\right]^2}{\sum_{j=1}^{N}\left[(Y)_{obs,j} - \overline{(Y)}_{obs,j}\right]^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left((Y)_{obs,j} - (Y)_{com,j}\right)^2} \quad (7)$$

RESULTS AND DISCUSSION

This paper studied the implementation of ANN and MLR models in the simulation of different absorbance as the out parameters using different concentrations as its corresponding input parameters. The modelling procedure for ANN and MLR were carried out in MATLAB (2017) and E-view software (2018), respectively. The simulation results based on the two widely used performance criteria are presented in Table 1. For the purpose of this paper ANN-A1 to ANN-A3 stands for simulation result for absorbance 1 to absorbance 3 derived from the peak area of the experimental result in the chromatogram, likewise in the case of MLR-A1 to MLR-A3.

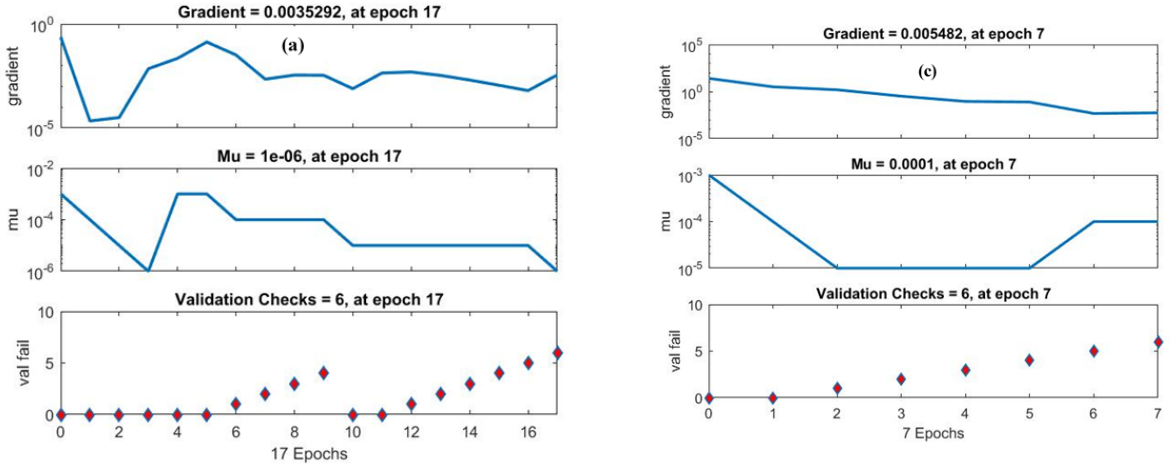
Table 1 Results of the ANN and MLR models

	Training		Testing	
Models	R^2	RMSE	R^2	RMSE
ANN-A1	0.9954	0.0201	0.9900	0.0236
ANN-A2	0.9992	0.0097	0.9997	0.0150
ANN-A3	0.9993	0.1762	0.9993	0.1983
MLR-A1	0.98983	0.1072	0.98890	0.1668
MLR-A2	0.9824	0.0446	0.97369	0.0447
MLR-A3	0.9577	0.8389	0.9538	0.8998

The results from Table 1 shows the predictive ability of both the linear and non-linear models for the simulation of absorbance based on different concentration(0.5, 1, 10, 25, 50, 100, 250, 500, 750, 1000ppm).RMSE satisfies the triangle inequality that is required for a distance function metric for model evaluation and it is preferred in data assimilation field where the sum of squared errors is often defined as the cost function to be minimized by adjusting model parameters. RMSE indicates the average magnitude of the model error and taking the relative mean squarevalue avoids error compensation. R^2 describes the degree of linear association between observed and predicted values directly such that it lies between 0 and 1 (with 1being a perfect model). Similarly, Elkiran et al., 2019 stated that for a good model simulation the performance indicator should include at least one goodness-of-fit (R^2) and one absolute error (RMSE).

Comparison for ANN and MLR models

Fig. 2 show the simulated pattern of the training function of ANN model.



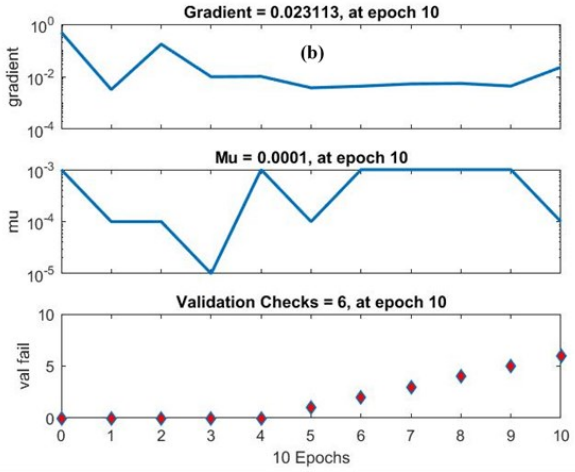
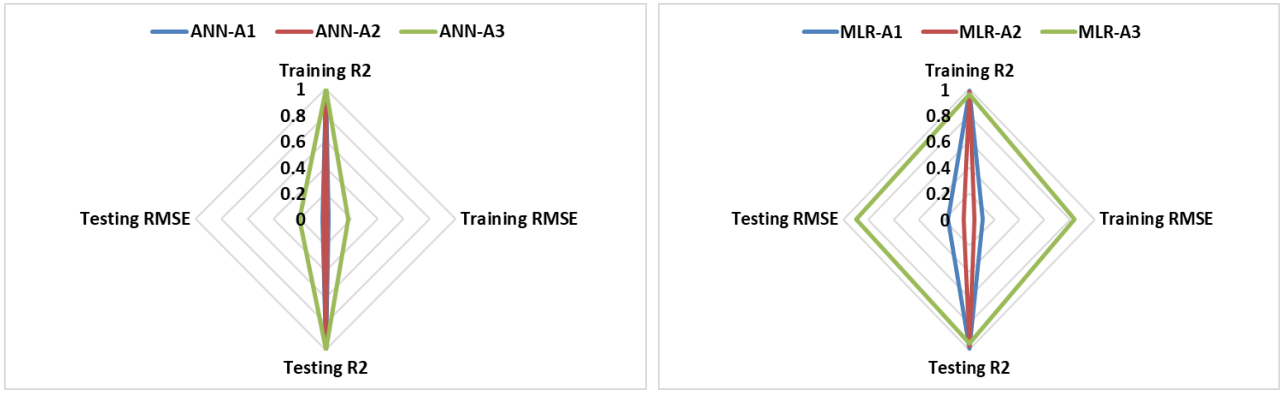


Fig. 2 Training function of the simulated ANN model for a) A1, b) A2 and c) A3

From Fig. 2 it can be seen that the model uses the trainlm - Levenberg–Marquardt function for training. Learngdm (gradient descent with momentum weight and bias learning function) function was used as the adaption learning function with different epoch (iteration) number. The comparison of the results for the two models demonstrated that for the simulation and assessing the absorbance, ANN model is more satisfactory and reliable. This conclusion was drawn by considering the values of R^2 and RMSE in both training and testing phase. The predictive skills of ANN model may be linked with its ability to handle complex and highly complex processes. As such this mechanism was the primary and basic reason why ANN have higher performance as compared with the classical linear regression MLR. Fig. 3 show the radar chart for the predictive models showing the comparison analysis between the training and testing phase.



(a) (b)

Fig. 3 Radar chart showing the R^2 and RMSE level for a) ANN and b) MLR models

The analysis of the results can equally be presented in a 3D surface diagram (Fig. 4). Based on the illustrative presentation itcan be seen that MLR model also provided the good predictive results even though it is a linear model. The conclusion could be drawn that the simulation of absorbance with different concentration is not highly stochastic processes due to the ability of the linear model to cope with the process, this is in line with (Arabameri et al., 2019). Moreover, the general results proved the ability of the both linear mathematical and nonlinear artificial intelligence model for the simulation of absorbance concentrations, with superiority of nonlinear in term of predictive skill of the model.

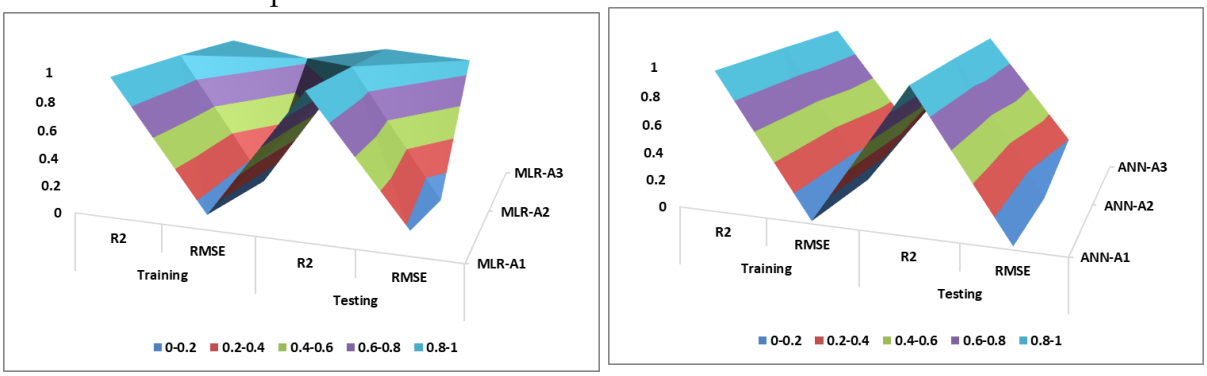


Fig. 4 Surface diagram showing the R^2 and RMSE level for a) ANN and b) MLR models

CONCLUSION

In this paper, Artificial Neural Network (ANN) and Multilinear Regression (MLR) analysis models were used. The absorbance data were measured from experimental analysis using HPLC instrument. The obtained results showed ANN model slightly outperformed MLR model for the simulation of all the three different absorbance. The predictive results also demonstrated that both the two models are capable of modelling the absorbance based on the performance efficiency of the model. Finally, the outcome suggested the used of other models such as adaptive neuro-fuzzy model (ANFIS), support vector machine (SVM) etc. in order to compare and come up with the best model and enhance the performance of the ANN and MLR prediction.It is equally recommended and suggested to try the applications of other nonlinear data-driven models and optimization algorithms such as support vector machine (SVM), principal component analysis (PCA), Adaptive neuro fuzzy inference system (ANFIS) fuzzy logic (FL), genetic algorithms (GA), etc. in order to improve the model performance accuracy.

REFERENCES

Abba, S.I. and Elkiran, G., 2017. Effluent prediction of chemical oxygen demand from the astewater treatment plant using artificial neural network application. *Procedia computer science*, *120*, pp.156-163.

Abba, S.I., Pham, Q.B., Usman, A.G., Linh, N.T.T., Aliyu, D.S., Nguyen, Q. and Bach, Q.V., 2020. Emerging evolutionary algorithm integrated with kernel principal component analysis for modeling the performance of a water treatment plant. *Journal of Water Process Engineering*, *33*, p.101081.

Abba, S.I., Usman, A.G. and Selin, I.Ş.I.K., 2020. Simulation for response surface in the HPLC optimization method development using artificial intelligence models: A data-driven approach. *Chemometrics and Intelligent Laboratory Systems*, p.104007.

Agatonovic-Kustrin, S., Zecevic, M., Zivanovic, L.J. and Tucker, I.G., 1998. Application of artificial neural networks in HPLC method development. *Journal of pharmaceutical and biomedical analysis*, *17*(1), pp.69-76.

Arabameri, M., Nazari, R.R., Abdolshahi, A., Abdollahzadeh, M., Mirzamohammadi, S., Shariatifar, N., Barba, F.J. and Mousavi Khaneghah, A., 2019. Oxidative stability of virgin olive oil: evaluation and prediction with an adaptive neuro-fuzzy inference system (ANFIS). *Journal of the Science of Food and Agriculture*, *99*(12), pp.5358-5367.

Chang, Q., Zuo, Z., Chow, M.S. and Ho, W.K., 2005. Difference in absorption of the two structurally similar flavonoid glycosides, hyperoside and isoquercitrin, in rats. *European Journal of Pharmaceutics and Biopharmaceutics*, *59*(3), pp.549-555.

Chen, W.B. and Liu, W.C., 2015. Water quality modeling in reservoirs using multivariate linear regression and two neural network models. *Advances in Artificial Neural Systems*, *2015*.

Elkiran, G., Nourani, V., Abba, S.I. and Abdullahi, J., 2018. Artificial intelligence-based approaches for multi-station modelling of dissolve oxygen in river. *Global Journal of Environmental Science and Management*, *4*(4), pp.439-450.

Gaya, M.S., Zango, M.U., Yusuf, L.A., Mustapha, M., Muhammad, B., Sani, A., Tijjani, A., Wahab, N.A. and Khairi, M.T.M., 2017. Estimation of turbidity in water treatment plant using Hammerstein-Wiener and neural network technique. *Indonesian Journal of Electrical Engineering and Computer Science*, *5*(3), pp.666-672.

Khademi, F. and Behfarnia, K., 2016. Evaluation of concrete compressive strength using artificial neural network and multiple linear regression models.

Nourani, V., Elkiran, G., Abdullahi, J. and Tahsin, A., 2019. Multi-region modeling of daily global solar radiation with artificial intelligence ensemble. *Natural Resources Research*, *28*(4), pp.1217-1238.

Romani, A., Galardi, C., Pinelli, P., Mulinacci, N. and Heimler, D.H.P.L.C., 2002. HPLC quantification of flavonoids and biflavonoidsinCupressaceae leaves. *Chromatographia*, *56*(7-8), pp.469-474.

Ruggieri, F., D'Archivio, A.A., Carlucci, G. and Mazzeo, P., 2005. Application of artificial neural networks for prediction of retention factors of triazine herbicides in reversed-phase liquid chromatography. *Journal of Chromatography A*, *1076*(1-2), pp.163-169.

You, H., Ireland, B., Moeszinger, M., Zhang, H., Snow, L., Krepich, S. and Takagawa, V., 2019. Determination of bioactive nonvolatile ginger constituents in dietary supplements by a rapid and economic HPLC method: Analytical method development and single-laboratory validation. *Talanta*, *194*, pp.795-802.