

# Potential of Hybrid Adaptive Neuro Fuzzy Model in Simulating Clostridium Difficile Infection Status



Ahmed Nouri Alsharksi, Y.A Danmaraya, Hadiza Usman Abdullahi, Umar Muhammad Ghali, A.G Usman

**Abstract:** The global burden posed by nosocomial diarrhea lead to the strong given attention by health practitioners science its morbidity and mortality rate hit about 500,000 rates annually in the United states. Diagnostic measures have been put in place to detect the presence of CD using different methods. Reliable prediction of the health status of patients is of paramount importance. This study aimed at investigating the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient from the record units of Near East University Hospital using hybrid adaptive neuro fuzzy (known as ANFIS) model consisting of various combinations of membership functions and training Fis. In this context, the age of the patients, gender, results of the analysis conducted, the department in which the patient was admitted, the age category and the hospitalization were employed as the input variables. The performance accuracy of these membership functions and training FIS combinations were checked using two performance indices determination coefficient (R<sup>2</sup>) and mean square error (MSE). The obtained computation data driven models proves the reliability of the combination of subtractive clustering membership function and hybrid training FIS over the other three ANFIS combinations. Overall, the results indicated the reliability and satisfaction of hybrid adaptive neuro fuzzy in checking the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient.

**Keywords:** Clostridium difficile; Stool samples; Hybrid adaptive neuro fuzzy; Membership functions; Training FIS.

## I. INTRODUCTION

Clostridium difficile (CD) is a nosocomial bacteria that is known for the release of cytotoxin A and B due to microbiome disruption. CD is mostly associated with health care, but community based CDs are also found due to the excessive use of unrecommended antibiotics. Several risk factors such as toxin A production, age, gastrointestinal procedures, renal insufficiency, are associated with health-associated CDI (Guh & Kutty, 2018). Its transmission

is possible as a result of different genetic composition of every patient, thereby different outcomes (Evans & Safdar, 2015). Clinical outcomes tell about the health status of the evaluated sample either positive (diseased) or negative (normal). These clinical outcomes reveal a high correlation between CDI markers and intestinal inflammation markers (Feghaly et al., 2013). The most common treatment available in the clinical guideline is orally administered vancomycin. Alternative treatment such as modeling of immunoregulatory therapeutics was employed to show the efficacy of lanthionine synthetase C-like (Feghaly et al., 2013).

Many diagnostic procedures exist for testing CD, but the toxigenic culture method is regarded as the gold standard. Maeda et al (2019) developed a computer-aided system that predicts the extent of histologic ulcerative inflammation using endoscopy. This computer-aided diagnostic give a sensitivity of 74%, specificity of 97%, and accuracy of 91% with a perfect reproducibility ( $\kappa=1$ ) [1].

Li et al (2019) recently reported a machine learning approach to predict C. difficile complications using electronic health record. This will surely help clinical studies in investigating the prevalence of this infection. Another study by Marra et al (2020) also shows the application of machine learning in predicting the presence of C. difficile in hospitalized patients. Looking at the aforementioned studies, it is now clear that the pivotal role played by artificial intelligence (AI) in different fields especially in the clinical aspect cannot be undermined. Therefore, this study aimed at investigating the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient from the record units of Near East University Hospital [2].

In this research, a hybrid adaptive neuro fuzzy (known as ANFIS) model consisting of various combinations of membership functions and training Fis was employed for simulating the health status of various subjects both inpatients and outpatients.

## II. MATERIALS METHODS

Patients (in-patients and outpatients) details were obtained (n = 230) for C. difficile related diarrhoea between the period 1st September 2015 to 31st December 2018. The data record was obtained from the stored records in different units of Near East University Hospital, Nicosia, North Cyprus. Stool specimens were collected from patients (both in-patients and out-patients) in a non-sterile container from different hospital units with the consent of suspected patients.

Revised Manuscript Received on July 20, 2020.

\* Correspondence Author

**A.G Usman\***, Faculty of Pharmacy, Department of Analytical Chemistry, Near East University, Nicosia, Turkish Republic of Northern Cyprus. E-mail: abdullahigusman@gmail.com

**Ahmed Nouri Alsharksi**, Department of Biochemistry, faculty of Medicine, Misurata University, Libya.

**Y.A Danmaraya**, Department of Chemistry, Faculty of Science, Yusuf Maitama Sule University, Ado Bayero house, Kano, Nigeria.

**Hadiza Usman Abdullahi**, Department of Chemistry, Faculty of Science, Federal University Dutse, Jigawa, Nigeria.

**Umar Muhammad Ghali**, Faculty of Medicine, Department of Medical biochemistry, Near East University, Nicosia, Turkish Republic of Northern Cyprus.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Samples were then stored at 40C prior to analysis. Mini VIDAS (Biomérieux; Serial No.: IVD 1206079) was used to detect toxins A and B according to the manufacturer’s procedure.

2.1 Proposed methodology

For predicting the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient from the record units of Near East University Hospital. In this study, the Adaptive neuro-fuzzy inference system (ANFIS) was proposed by using a combination of different membership functions (Subtractive clustering (SC) and Grid partitioning (GP)) and training FIS (back propagation (back) and hybrid) . The data involved in this study were taken from record units of Near East University Hospital. For this, the health status of various subjects were predicted using six different input parameters including the age of the patients, gender, results of the analysis conducted, the department in which the patient was admitted, the age category and the hospitalization. Some of these parameters were denoted in the form of numerical values while some of them were converted into numerical values. In the process of the models development, Fig 1. depicted the flow chart of data driven models as well as the experimental procedure involved in these work.

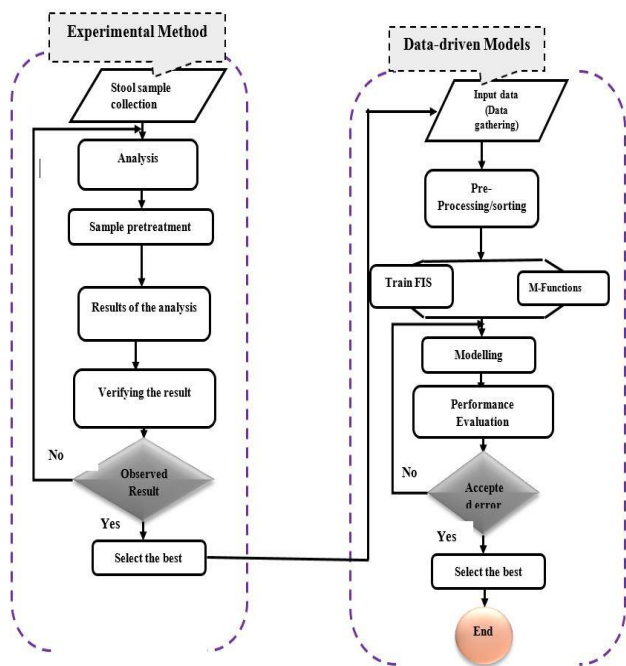


Fig 1. Summary of the experimental and the data driven method

2.2 Adaptive neuro-fuzzy inference system (ANFIS)

ANNs tools are one of the broadly use AI-based model which are 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 [3].

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 [4]. ANFIS is a real-world estimator because of its capacity to approximate real functions [5]. In practice, several membership functions are used including

trapezoidal, triangular, sigmoid and Gaussian. Although the Gaussian function is the most frequent MF.

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  $\mu(x)$  is  $A_1$  and  $\mu(y)$  is  $B_1$  then  $f_1 = p_1x + q_1y + r_1$  (1)

Rule 2: if  $\mu(x)$  is  $A_2$  and  $\mu(y)$  is  $B_2$  then  $f_2 = p_2x + q_2y + r_2$  (2)

$A_1, B_1, A_2, B_2$  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. Refer to (Abba et al., 2019) for more information about ANFIS.

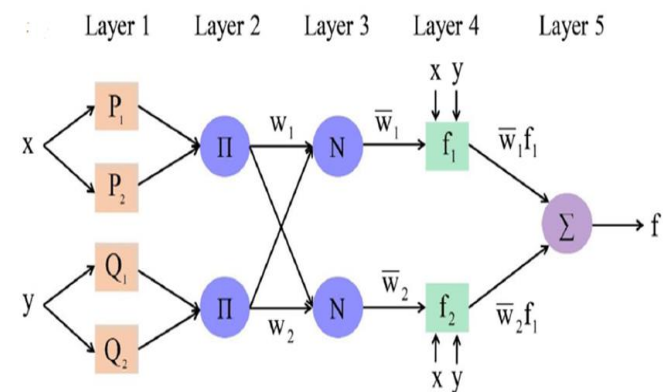


Fig. 2: Schematic structure of ANFIS model [3].

2.2.1 Grid Partitioning (GP)

This is a form of artificial intelligence that combines both grid partitioning and ANFIS. In Grid partition technique, the data is splitted into grid data, according to the number and kind of the membership functions (MFs), which is present in each of the dimensions [6]. The GP mechanism starts with zero output and then continuously keep learning the various fuzzy rules set as well as functions via the calibration method [7]. Least square method is applied in order to determine the initial fuzzy sets and variables, which depends on the MFs as well as the partition types by employing and optimizing two different training FIS know as hybrid and back propagation [8].

2.2.2 Subtractive Clustering (SC)

The subtractive clustering is another class of model that involves the combination of clustering and ANFIS approach. For modelling SC, each data point is regarded as a possible cluster center [9]. Thus, any point, which is composed of many points neighbouring it gives a higher possible value [10]. Therefore, to find the first cluster centre, the definition of the measure of density (di) is shown in Eq. 3. Whereby, the data point that composed of the highest potential value or density is considered as the first cluster centre [11].

$$d_i = \sum_{k=1}^N \exp\left(-\left(\frac{2}{r_o}\right)^2 \cdot \|x_i - x_k\|^2\right) \quad (3)$$



2.3 Validation and performance Evaluation methods

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 [12]. In this technique, the initial data set is categorized into same-sized subsets of k [13]. Regarding the performance evaluation of the models, two performance criterias are considered, which composed of determination coefficient (R2) as a goodness of fitness and Mean Squared Error (MSE) to accounts for the errors depicted by models developed using a combination of various membership functions and training FIS for the method development [13].

$$R2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \tag{4}$$

$$MSE = \frac{\sum(y_i - \hat{y}_i)^2}{N} \tag{5}$$

$$i = 1, 2, 3 \dots \dots \dots N$$

Where  $y_i, \hat{y}_i, \bar{y}_i$  and N stands for actual data, simulated data, Mean value of the actual data and total number of data instances respectively [14].

III. RESULTS AND DISCUSSION OF THE COMPUTATIONAL METHOD

In this study, the application of ANFIS as well as the comparative analysis for the combination of its membership function (Grid partitioning (GP) and subtractive clustering (SC)) and its training FIS (Back propagation and hybrid) was explored to predict the health status of patients and subjects, using six different input parameters including the age of the patients, gender, results of the analysis conducted, the department in which the patient was admitted, the age category and the hospitalization. The linearity of the parameter from the raw data was checked through the application of the Pearson correlation. As shown in table 1, there is a strong inverse correlation between the status of the patients with the results of the analysis conducted on each patient with R-value= -0.4679. More also, there is an average direct correlation between the age category of the patients with the result of the analysis having R= 0.1239. Table 1 further describes the statistical values of the data involved in this study.

Table 1: Statistical and Correlation analysis Statistical analysis

Parameters	Result	Age	Gender	Department	Age category	Hospitalization	Status
Mean	0.2522	0.465913	0.15304	8.1391	2.6913	1.5522	1.8391
Median	0.0400	0.450000	0.0000	8.0000	3.0000	2.0000	2.0000
Mode	0.0200	0.210000	0.0000	8.0000	2.0000	2.0000	2.0000
Standard Deviation	1.0310	0.217976	0.5002	2.6839	0.9230	0.4984	0.3682
Kurtosis	39.8372	-1.0545	-2.0024	0.8167	-1.0877	-1.9726	1.4656
Skewness	6.2232	0.0836	-0.1228	0.3493	0.1481	-0.2112	-1.8582
Minimum	0.0100	0.0000	0.0000	1.0000	1.0000	1.0000	1.0000
Maximum	8.1500	0.910000	2.0000	15.0000	4.0000	2.0000	2.0000

Correlation Analysis

Parameter	Result	Age	Gender	Department	Age category	Hospitalization	Status
Result	1.0000						
Age	0.0731	1.0000					
Gender	0.0121	0.2030	1.0000				
Department	-0.1054	-0.1985	0.0229	1.0000			
Age category	0.1239	0.9500	0.1765	1.0000	1.0000		
Hospitalization	-0.0680	-0.3277	-0.0239	2.0000	-0.0640	1.0000	
Status	-0.4680	0.0451	0.0148	4.0000	-0.0054	-0.1563	1.0000

The performance accuracy of the models was evaluated using two different performance indices: determination coefficient (R<sup>2</sup>) as a goodness of fitness and mean square error (MSE). In the development of these models, the simulation was done in MATLAB 9.3 (R2019a).

In modelling ANFIS, different kinds of membership functions as well as epoch iterations were employed through trial by error in order to determine the desired structure [15]. The best architecture of the model was optimized and selected using trial by error method Tables 2 shows the results of the performance of the computation method in modelling the status of the subjects.

Table 2: Results of the computational data driven models

	Calibration		Verification	
	R2	MSE	R2	MSE
Grid-Hybrid	0.9932	0.0011	0.9796	0.0024
Grid-Back	0.9487	0.0060	0.9400	0.0101
Sub-Hybrid	0.9998	0.0004	0.9992	0.0009
Sub-Back	0.9521	0.0056	0.9135	0.0145

The performance accuracy of the computational data driven models is shown in table 2. Based on the result of the models. It can be seen that all the four models that composed of different membership functions (Grid partitioning and Subtractive clustering) and two different training FIS optimization methods (Back propagation and hybrid) of ANFIS are capable of simulating the status of the subjects. Nevertheless, the combination of subtractive clustering membership function and hybrid training FIS optimization method (Sub-Hybrid) proves to have shown higher efficiency as compared with the other combinations of the ANFIS model in both the calibration and verification phases. Furthermore, the performance efficiency of the models in terms of R2 depicted that Sub-Hybrid outperformed the other three model combinations Grid-Hybrid, Grid-Back and Sub-Back in the verification stages and increase their performance skills approximately by 2%, 6% and 9% respectively.

3.1 Comparative performance of the computation algorithms

The results of the models can also be shown graphically using a scatter plot in order to comparatively prove table 2. It can be seen from Figure 3 that all the combinations of the ANFIS model shows a reliable and satisfactory result.





Nevertheless, the combination of Sub-Hybrid ANFIS model proves to be rugged and robust as compared with the other three ANFIS model combinations. This is in line with table 2. Further comparative analysis of the scatter plot proves a higher agreement between the measured and

predicted values were associated more to the Sub-Hybrid combination as compared with the other three combinations of ANFIS model.

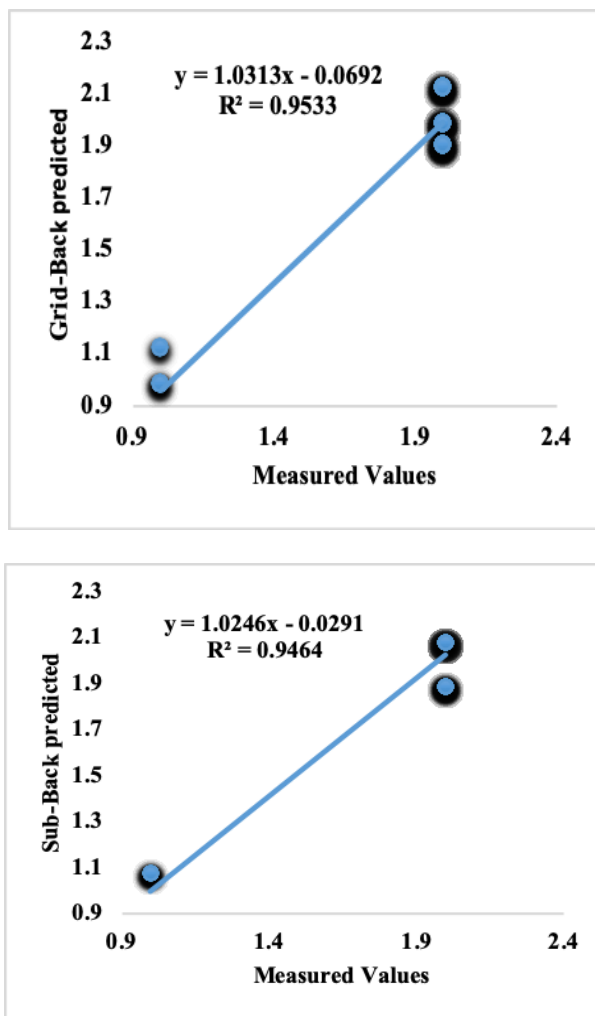


Fig. 3: Scatter plots of the computational data driven models for modelling the *clostridium difficile* infection status

The comparative performance of these computational models can equally be demonstrated using a bar chart in terms of MSE. Whereby, the lesser the MSE the better the performance of the model. Figure 4 below indicated that Sub-Hybrid proved to be the best performing model in terms of MSE. This findings is in line with other studies in the technical literature such as [16],[17], which demonstrated an error values higher than 0.04.

The comparative predictive skills of the models is also depicted graphically using a radar chart. Figure 5, was employed in order to comparatively shows the correlation co-efficient of each combination of the models, which is aimed to show the multi variable quantitative analysis.

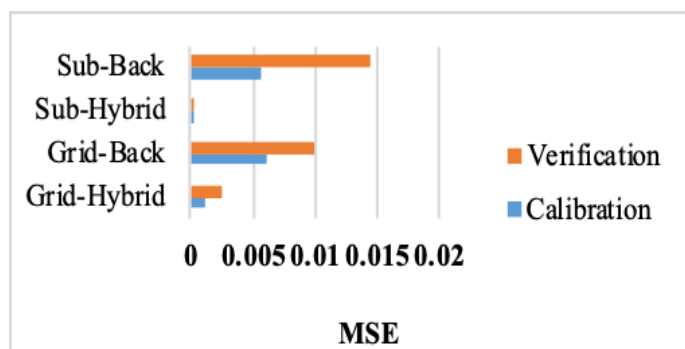


Fig.4: Comparative performance of the computational model in terms of Mean squared Error

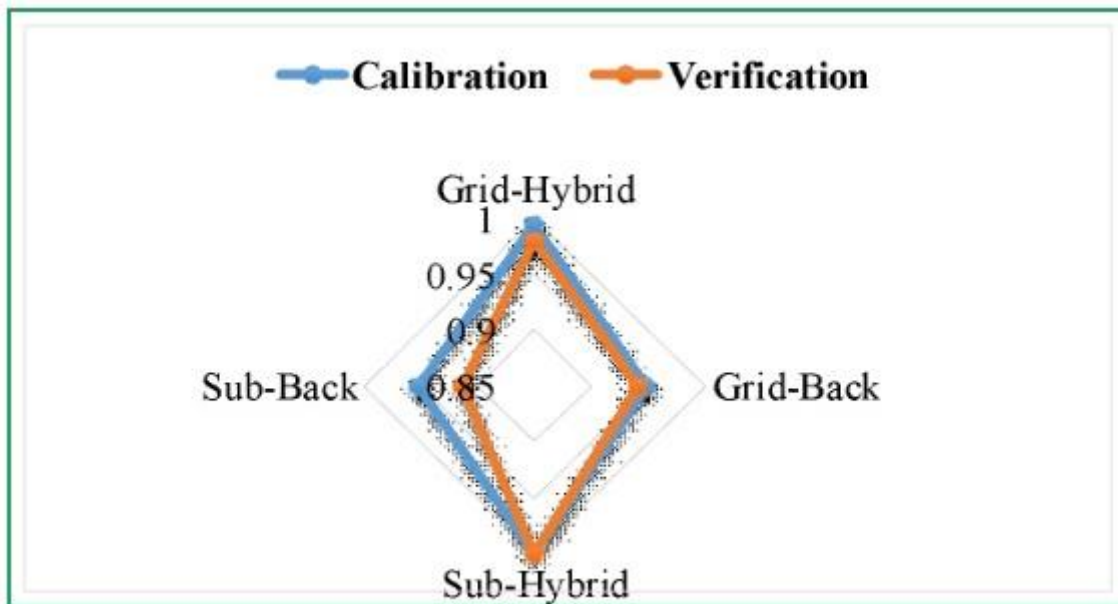


Fig. 5: Predictive comparison of the models in terms of correlation co-efficient using radar chart

#### IV. CONCLUSION

The global burden posed by nosocomial diarrhea lead to the strong given attention by health practitioners science its morbidity and mortality rate hit about 500,000 rates annually in the United states. Therefore, diagnostic measures have been put in place to detect the presence of CD using different methods. Therefore, the development of a robust computational intelligence system is of paramount importance for checking the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient. In this study, the data was collected from the record units of Near East University Hospital. The data was further simulated using hybrid adaptive neuro fuzzy (known as ANFIS) model consisting of various combinations of membership functions and training Fis. The prediction efficiency of these model combinations were analysed and enhanced using two statistical measures including determination coefficient (R2) and Mean squared error (MSE) as well as some graphical plots such as scatter plots in order to demonstrate the fitness between the experimental values and the predicted values, radar and bar chart, which are used to in order to indicated the predictive comparative performance of the models. Furthermore, the comparative evaluation of the ANFIS model combinations shows that subtractive clustering membership function in combination with hybrid train FIS (Sub-hybrid) outperformed all the other two combinations and increase their performance accuracy. In general, this study indicated the reliability of ANFIS model in modelling the status of stool samples collected to test the presence of clostridium difficile as either positive or negative from both inpatient and outpatient. The study equally recommends that other non-linear models such as support vector machines (SVM) and Hammestein weiner (HW) can be employed to model the performance of the status of patients tested with clostridium difficile as either positive or negative from both inpatient and outpatient.

#### REFERENCES

1. Y. Maeda, S. Kudo, Y. Mori, and M. Misawa, "Fully automated diagnostic system with arti fi cial intelligence using endocytoscopy to

identify the presence of histologic in fl ammation associated with ulcerative colitis ( with video )," *Gastrointest. Endosc.*, vol. 89, no. 2, pp. 408–415, 2019.

2. B. Y. Li, J. Oh, V. B. Young, K. Rao, and J. Wiens, "Using Machine Learning and the Electronic Health Record to Predict Complicated Clostridium difficile Infection," pp. 1–7, 2013.

3. H. U. Abdullahi, A. G. Usman, and S. I. Abba, "Modelling the Absorbance of a Bioactive Compound in HPLC Method using Artificial Neural Network and Multilinear Regression Methods," vol. 6, no. 2, pp. 362–371, 2020.

4. U. M. Ghali *et al.*, "Applications of Artificial Intelligence-Based Models and Multi- Linear Regression for the Prediction of Thyroid Stimulating Hormone Level in the Human Body," vol. 29, no. 4, pp. 3690–3699, 2020.

5. S. I. Abba, A. G. Usman, and S. İşik, "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.

6. M. S. Gaya, N. A. Wahab, Y. M. Sam, A. N. Anuar, and S. I. Samsuddin, "ANFIS modelling of carbon removal in domestic wastewater treatment plant," *Appl. Mech. Mater.*, vol. 372, pp. 597–601, 2013.

7. 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.

8. V. Nourani, S. Uzelaltinbulat, F. Sadikoglu, and N. Behfar, "Artificial intelligence based ensemble modeling for multi-station prediction of precipitation," *Atmosphere (Basel)*, vol. 10, no. 2, pp. 1–28, 2019.

9. G. Elkiran, V. Nourani, and S. I. Abba, "Multi-step ahead modelling of river water quality parameters using ensemble artificial intelligence-based approach," *J. Hydrol.*, vol. 577, no. April, p. 123962, 2019.

10. S. M. Ghoreishi, A. Hedayati, and S. O. Mousavi, "Quercetin extraction from Rosa damascena Mill via supercritical CO<sub>2</sub>: Neural network and adaptive neuro fuzzy interface system modeling and response surface optimization," *J. Supercrit. Fluids*, vol. 112, pp. 57–66, 2016.

11. M. Quiñones-Grueiro, A. Prieto-Moreno, C. Verde, and O. Llanes-Santiago, "Data-driven monitoring of multimode continuous processes: A review," *Chemom. Intell. Lab. Syst.*, vol. 189, pp. 56–71, 2019.

12. Q. B. Pham, S. I. Abba, A. G. Usman, N. Thi, and T. Linh, "Potential of Hybrid Data-Intelligence Algorithms for Multi-Station Modelling of Rainfall," 2019.

13. A. G. U. Selin and I. S. I. Abba, "A Novel Multi - model Data - Driven Ensemble Technique for the Prediction of Retention Factor in HPLC Method Development," *Chromatographia*, no. 0123456789, 2020.
14. 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, 2020.
15. 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.
16. 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.
17. 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.

### AUTHORS PROFILE

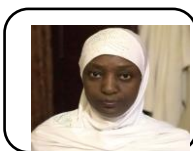


**Ahmad Nouri Alsharki** is a Certified Medical Lab Technician with degree in medical laboratory and a Certified master degree in medical and clinical Microbiology department with 10 years of experience in patient testing, result evaluation and instrument maintenance and troubleshooting. Well-versed in collecting, handling and processing specimens and reporting and

maintaining test results. Medical Laboratory Technician skilled in collecting and processing biological specimens for laboratory testing and analysis. Strong knowledge of laboratory Instrumentation, procedural calibration and quality control review, assessment and subsequent reporting according to laboratory protocol and procedures. Published one paper titled Applications of artificial intelligence-based models and multi linear regression for the prediction of thyroid stimulating hormone level in the human body at International Journal of Advanced Science and Technology.



**Yahya Aliyu Dan Maraya** is a certified chemist, who bagged his first degree from Bayero University Kano and Master Degree from Near East University TRNC. He is currently working as a lecturer at Yusuf Maitama Sule University Kano, department of chemistry.



**Hadiza Usman Abdullahi** received her B.Sc. (Hons) in Chemistry from Kano University of Science and Technology (KUST) Wudil, Nigeria in 2012 and her M.Sc. in Inorganic chemistry from Jodhpur National University, India in 2016. Her research interest includes synthesis and characterization of inorganic complexes, water treatment and quality control. Currently working with Department of chemistry Federal University Dutse, Nigeria.



**Umar Muhammad Ghali** is a certified Medical Biochemist with a B Sc in Biochemistry from Al-Hikmah University Ilorin, Kwara State, Nigeria. He obtained a master degree in Medical Biochemistry from near East University, TRNC. His research interest includes; Cancer research, Phytochemistry, chemical kinetics, chromatography, toxicology, enzymes and

applied artificial intelligence.



**Abdullahi Garba Usman** bagged his bachelor degree from Bayero University, Kano (BUK), master's degree from Near East University TRNC from the department of pharmaceutical chemistry. Currently he is a Ph.D. student at Near East University, TRNC in the department of analytical chemistry. His current research interest includes chemometrics, phytochemistry, water resources, chromatographic modeling and simulation, water quality and wastewater treatment plant. He also has additional skills and expertise in data-driven algorithms, Artificial intelligence and data processing. He has numerous publications on chromatographic modelling and simulations. His recent publication is entitled 'A Novel Multi-model