Prediction of Artesian Source Depollution in Rainy Season Using an Artificial Neural Network

Diarra Mamadou, Tiendrebeogo B. Telesphore, Séré Abdoulaye and Sanogo Sy Rachid

December 1, 2023

ISSN 2442-6571

Prediction of Artesian Source Depollution in Rainy Season using an Artificial Neural Network

DIARRA Mamadou[a,1,*], TIENDREBEOGO B. Telesphore[b,2], SERE Abdoulaye[b,3], SANOGO Sy Rachid[a,4]

[a] Polytechnic School of Ouagadougou (EPO), Ouagadougou, Burkina Faso
[b] Nazi Boni University (UNB), Bobo-Dioulasso, Burkina Faso
[1] diarra.md@epo.gov.bf, *; [2] tetiendreb@gmail.com; [3] abdoulaye.sere@recifaso.org, syrachidsanogo78@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received
Revised
Accepted

Keywords
Water pollutants
Big Data
Artificial Neural Networks
Groundwater

ABSTRACT

The management of water suitable for human consumption is a major concern in urban and peri-urban areas. Water reservoirs are regularly polluted as a result of intense and non-localized rainfall events. Predicting these rapid and complex events is very difficult as the drainage of runoff and wastewater is often poorly mastered with massive manpower, time, cost and data.

An artificial neural network (ANN) is an effective method for time series prediction, analysis and forecasting Big data in various science and engineering disciplines. It is also beginning to be used in water management.

This study analyzes the ability of multilayer perceptron ANNs to inhibit the pollution of an artesian source in contact with runoff from the city of Bobo-Dioulasso. The results show that ANNs offer a better performance in terms of water quality prediction of an artesian source.

1. Introduction

Water covers 72 % of the surface of the globe, with an estimated volume of 1400 million km3. This volume is 97.2 % salty and is found in the oceans, inland seas, but also in some groundwater. Fresh water represents only 2.8% of the total water of the globe, made up of 2.1% of ice and snow and 0.7% of available fresh water [1]. Half of this 0.7% is groundwater [2]. Population growth, urban development and global warming are all weighing heavily on the management of water resources, leading to ever-changing needs [1].

Thus, a sustainable management of freshwater requires a global reflection that takes into account all the problems, constraints and challenges for its use. The depollution of rainwater and wastewater is an imperative for the preservation of the water cycle, insofar as they can pollute the water table at any time [3]. Therefore, maintaining water quality and making it safe for consumption becomes an absolute necessity. Water production requires laboratory analysis of different pollution parameters with durations that vary from a few hours to several days depending on the pollution level [4]. However, advances in Big data mining, complex pattern recognition, prediction of complex variables and machine learning algorithms allow to analyze data continuously, to automate routine and critical decisions without delaying human judgment.

This study allows the elaboration of a knowledge base for the depollution of artesian sources. A triple gain is expected :

1) Time saving in the construction of knowledge bases of different pollutants that make artesian sources unfit for consumption with a saving in the time of availability and a reduction in the cost of production;

2) qualitative advantage in the knowledge base construction of the inhibitors of the different hydric pollutants, thus favoring the reuse of the artesian sources by a faster decision making;

3) organizational and strategic gain in the multi-expert differentiation and management of the convergence between pollutants and their inhibitors.

This document is structured as follows : After an introduction where we will give the objectives of the paper. In section 2, a related works of water and the impact of Big data on the hydroinformatics decision support is made. Section 3, is devoted to a contribution to the potabilization of an artesian source in real time using artificial neural networks. Finally, conclusions are drawn and future work is described in section 4.

2. Related Works

2.1. Generalities of waters

Water whose chemical denotation is H_2O is a vital substance, and water is signified with a chemical bond that is formed between two hydrogen elements and one oxygen element [6].

Rainwater and runoff are two facets of the same water that circulates under, on and through the city. Groundwater is a more or less deepwater table formed by the accumulation of infiltrations in the soil over time conditioned by the porosity and geological structure of the soil. Groundwater is usually sheltered from sources of pollution and is therefore of excellent physico-chemical and microbiological quality compared to surface water [7], [8].

An artesian source corresponds to a spontaneous gush of water through a natural orifice without the need for drilling to reach the water table. This type of source corresponds to a karstic fissure at the level of which the water is under pressure. Such a phenomenon exists when the pressure level of the underground water becomes greater than the distance to the surface of the earth [9].

2.2. Water Quality

Drinking water or water intended for human consumption, according to the World Health Organisation (WHO), is water that can be drunk or used for domestic and industrial purposes without risk to health [6]. It must meet certain qualitative and quantitative physico-chemical and microbiological criteria set by regulation, classically referred to as the ”standard of potability” [8]. The water potability standards adopted in Burkina Faso are supervised by the Office National de l’Eau et de l’Assainissement (ONEA) [7], [9].

Water pollution is an alteration in the quality and nature of water that makes its use unsafe and/or disrupts the aquatic ecosystem. It can affect surface water (rivers, water bodies) and groundwater. Groundwater quality is characterized by the parameters it contains, their quantity and their effect on aquatic ecosystems and human health. Pollutants in runoff or leaching from the atmosphere and urban surfaces on the one hand and erosion on the other [10].

The Water Quality Index (WQI) is a method of analysing overall water quality using a group of parameters that reduce large amounts of information to a single number. This method was originally proposed by Horton [11] and Brown et al [12]. To calculate this index, Horton proposed the first formula that takes into account all the parameters needed to determine surface water quality and reflects the composite influence of different parameters important for water quality assessment and management [13]. There are 63 water potability criteria grouped into physico-chemical, organoleptic, microbiological, undesirable substances and toxic substances [3], [6].

Different water treatment methods are combined depending on the initial water quality and the objectives. The main techniques used are physical, chemical and biological. Physical methods are related to clarification to remove suspended solids (SS), solid-liquid separation. Chemical methods are based on a chemical interaction between the pollutants to be treated and inhibitors that neutralizes the harmful effects of the pollutants. Biological methods, aerobic and anaerobic processes are used, to to degrade organic compounds [4].

2.3. Arsenic geochemistry

Arsenic (As) is a heavy metalloid with emanating concern of its environmental toxicology worldwide. It is released into the living environment via natural and anthropogenic sources (Fig.1) [14]. As is one of the most problematic natural contaminants in groundwater worldwide. In fractured bedrock aquifers, natural concentrations of arsenic in groundwater can exceed the World Health Organisation guideline of µg/l [15]. Groundwater containing As above permissible levels is a worldwide occurrence and many millions of people rely on such water for their daily drinking water needs, with potentially serious and chronic consequences on their health [15]. Exposure to arsenic through consumption of water from domestic wells can lead to a variety of health various health problems including keratoses and skin pigmentation as well as lung and bladder cancers and diabetes.

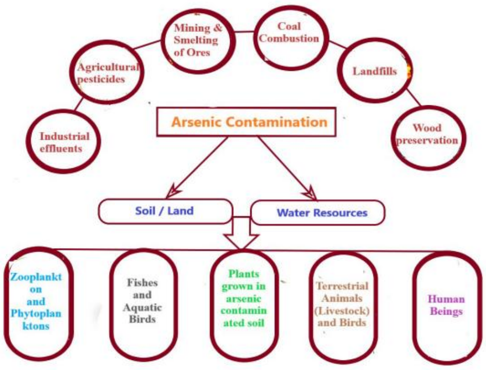

Fig. 1. Anthropogenic Sources of As contamination and the affected living organisms [14]

2.4. Big Data in Water Resources Engineering

Big data are characterised by five elements: volume, velocity, variety, veracity and veracity and value, defined as the five ”V” dimensions [16]. From the perspective of each characteristic, we can find data related to the water management problem [17].

Water treatment stations for human consumption generate a large volume of data on water quality. Effective analysis of all this data could result in competitive advantage and reduced decision uncertainty. Thus it is useful to apply Big data technology to the prediction of water quality and the treatment of water not fit for human consumption. Although, the level of adoption appears to be lower than in other sectors, where Big data and machine learning are now widely used to support evidence-based decision making [18].

The challenges of big data for water are high and the effectiveness of the proposed solutions is under increasing scrutiny. With Big data, we can model the content, quality and quantity of pollutants which are both complex and dynamic for water quality determination [19].

The increasing availability of water data has led to the emergence of predictive algorithms to modernise and stimulate water management. These models use data from a variety of sources such as the ideal measuring instruments used by water treatment plants under realistic conditions. Big data management depends on systems that can efficiently process and analyse large volumes of disparate and complex information. In this respect, Big data andartificial intelligence (AI) have a somewhat reciprocal relationship. Big data would be of little use without AI to organise and analyse it [500].

2.5. Artificial Intelligence for water management

Research in hydrology focuses on coupling deterministic and stochastic statistical modelling of processes to develop hydrological scenarios. Deterministic models are based on a known or hypothetical law of physics, mathematics or others disciplines, so that given input values always produce the same result. In contrast, the stochastic model accepts a certain probability distribution associated with given inputs, in the processes within the model and thus in the output, so that the same input can lead to different output values [4]. Therefore, there is a multitude of conceptual hydrological models, which differ either by the processes they take into account, or by the mathematical techniques used to solve the system of equations, or by the dimensionality of the problem (one, two or three dimensional; permanent or transient flow regime) [5].

The ISO 2382-28 standard defines artificial intelligence (AI) as ”the ability of a functional unit to perform functions typically associated with human intelligence, such as reasoning and learning”. Artificial Neural Networks (ANNs) should be considered as an improved AI method that is used to compute many multifaceted challenges in a rational time space [4]. A dynamic network can remember past information that is suitable for addressing complex dynamic and sequential problems. A nonlinear autoregressive network or loop network is a dynamic neural network with a high memory, whose inputs are exogenous variables and past values of the output, has been used to estimate groundwater quality (Fig. 2).

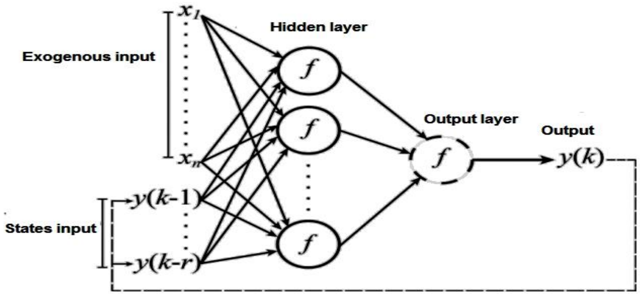

Fig. 2. A Dynamic Neural Network Model

The model outputs of the AI methods are also compared with predicted and observed values. Performance comparisons of the models are usually used the statistical expressions such as Mean Absolute Error (MAE), Mean Square Error (MSE), and Correlation Coefficient, scatter and time series graphs. Using Big data, business intelligence can develop predictive and prescriptive tools for water management. Advances in AI are opening the way to new alternatives for water management and treatment.

3. Results and Discussion

This study focused on ANN implementations for groundwater quality and treatment modelling. Our depollution model is made up of two phases illustrated by the : "Fig. 3" The first phase uses ONEA's potability criteria to initialise a reference database on the WQI. Then an algorithm calculates the risk quotient to evaluate the pollutant data of the collection. It uses the cross-correlation method, taking into account all the quality indicators identified. When it fails to classify a substance, it is directed to specialised sites in the cloud for identification and updating of the reference base. In the second. phase, an algorithm determines the quality index of the water after treatment by comparing the new content of pollutants with the drinking water standards. If not, an algorithm determines an exact measurement of the inhibitor to be used. When there is no When there is no listed inhibitor, a query is made on the cloud to update our reference database.

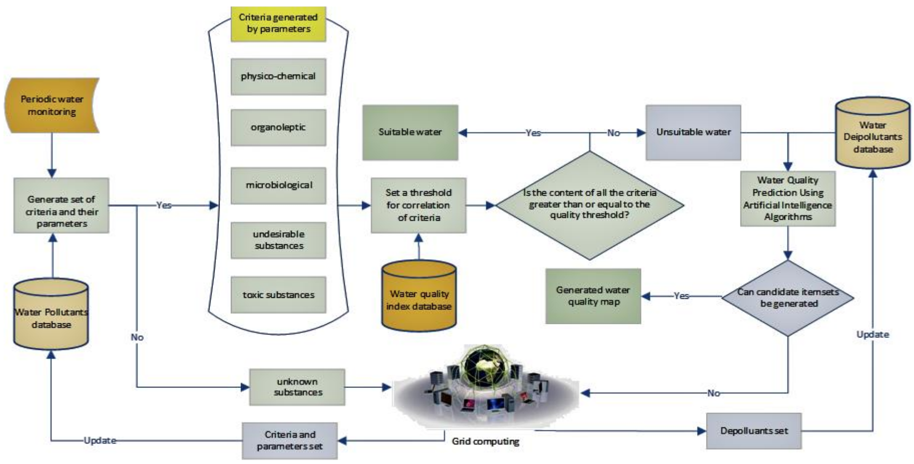

Fig. 3. Water quality prediction model

3.1. Model description

We evaluate the performance of the p-order autoregressive model AR (p) of the dynamic neural network, using prediction algorithms (algorithm 1), against traditional methods.

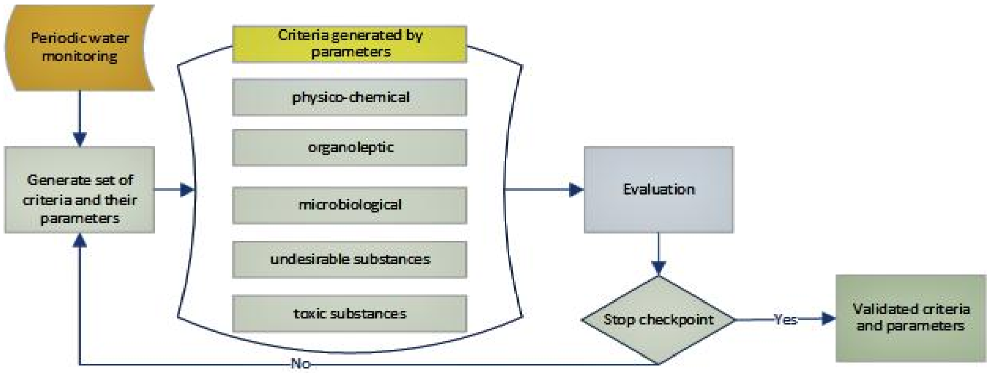

Fig. 4. Pollutants selection process

3.2. Prediction Algorithm Principles

The use of ANN requires the setting up of a learning base consisting of an input-output pair. This is a supervised model with a backpropagation algorithm that allows the optimal number of neurons in the hidden layer and the corresponding synaptic weights to be output. Prior to the training of our ANN models, the normalisation was applied for the data.

Normalisation is essential to ANN which makes the data dimensionless and confines them within a certain range. After training, the model that gives the best results in terms of Determination Coefficient (DC) (Eq. 1), Root Mean Squared Error (RMSE) (Eq. 2) is selected as the most efficient model. The linear correlation coefficient between the prediction and the observations is also measured by the r-square value (R2). DC measures the quality of a prediction, the closer this indicator is to 1, the closer the model is in reality. RMSE indicates the average difference between the predicted and observed values in a model. R returns the square of the Pearson correlation coefficient for two sets of values, which can be interpreted as the proportion of the variance in the prediction of the variance in the observation.

$$DC = R^2 = 1 - \frac{\sum_{i=1}^{N}(O_i - P_i)^2}{\sum_{i=1}^{N}(O_i - \bar{O})^2} \quad (1)$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{\sum_{i=1}^{N}(O_i - P_i)}{N}} \quad (2)$$

$$R = \frac{\sum_{i=1}^{N}(O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N}(O_i - \bar{O})^2 \times \sum_{i=1}^{N}(P_i - \bar{P})^2}} \quad (3)$$

Where P_i and O_i are respectively the predicted and observed values at time i, their mean and N the number of observed data.are prescribed.

3.3. Results and discussions

The normalisation of input data is an important step in data processing before the application of ANNs. Our ANN uses the collections of data on the rainy seasons of 2020 and 2021 from ONEA (Bobo-Dioulasso). For the estimation of the parameters of a water model, the data is divided into two parts. The first is used to calibrate the model and the second to validate it. This practice is known as split-sample testing. The size of the calibration data depends on the number of parameters to be estimated. The learning function is backpropagation. The training runs over 5000 cycles (a few hours on a 2.5 Ghz i5 core). We varied the number of nodes in our ANN from 5 to 20 in steps of 5, i.e. 5, 10, 15 and 20 nodes. We have a single output layer containing information on arsenic (As) which is a micropollutant of the toxic substance parameter. Its presence in contaminated water used for drinking, food preparation and irrigation constitute the greatest threat to public health [3].

We see in Table 1 that during calibration, the RMSE, R and DC values for all developed models vary from 1.57 to 2.35; 0.793 to 0.945 and 0.587 to 0.910 respectively. However, the values of RMSE, R and DC vary in the range of 1.56 to 3.82 cumec, 0.654 to 0.943 and 0.430 to 0.852 respectively during model validation. The best performing ANN model is the ANN15 model with RMSE, R, DC calibration values of 1.57, 0.945, 0.910 and validation values of 1.56, 0.943, 0.890

respectively. Our ANN models have as input variables micropollutants such as arsenic, cyanide, chromium, nickel, selenium and some hydrocarbons). The results indicate that the ANN model provides a better learning performance with an increase in the number of input variables. In the case of ANN20 with 20 input variables, the performance is poorer than that of ANN2 due to the higher number of input variables which increases the complexity of the model. Increasing the complexity of the model causes the model to overfit the training data, resulting in poor predictions.

In the case of ANN5 and ANN10, the performance decreases compared to ANN15 due to the reduced number of input variables from 15 to 10 and 10 to 5. Therefore, in developing an ANN model, it is very important to use an optimal number of input variables and for the present study, the results indicate that for the simulation of (As), the 15 input variables used in ANN15 are optimal. The comparative results of the simulated ANNs are shown below (Tab. 1).

Table 1. Comparative performance of various ANN models

ANN Models	Training (Calibration)			Testing (Validation)		
	RMSE	*R*	*DC*	*RMSE*	*R*	*DC*
ANN5	2.35	0.793	0.587	3.82	0.654	0.430
ANN10	1.89	0.840	0.687	2.05	0.794	0.629
ANN15	1.57	0.945	0.910	1.56	0.943	0.890
ANN20	1.71	0.815	0.753	1.85	0.928	0.862

Normalising input data is an important step in processing the data before applying ANNs. Our ANN uses data collected from the Nasso artesian spring, from ONEA (Bobo-Dioulasso), over the 2020 and 2021 rainy seasons. To estimate the parameters of a water model, the data is divided into two parts. The first is used to calibrate the model and the second to validate it. This practice is known as "split-sampling testing". The size of the calibration data depends on the number of parameters to be estimated. The learning function used is backpropagation (Algorithms 1, 2, 3 and 4). Our various ANN models use micropollutants such as arsenic, cyanide, chromium, nickel, selenium and certain hydrocarbons as input criteria. The results indicate that ANN models offer different learning performances with an increase in the number of input criteria.

Overall, the calibration of the different models, depending on the predictive and forecasting techniques, gives us an output that varies from 1.57 to 2.35 for the RMSE, from 0.793 to 0.945 for the R and from 0.587 to 0.910 for the DC (Fig. 5). Validation gave us an output of RMSE varying between 1.56 and 3.82, R between 0.654 and 0.943 and DC between 0.430 and 0.852 (Fig. 6).

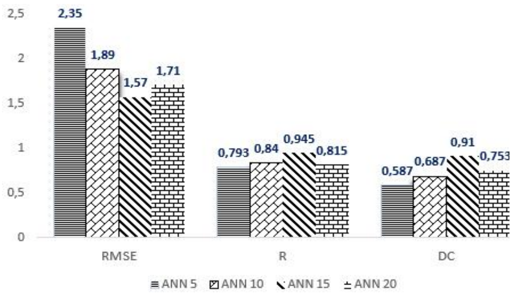

Fig. 5. Models Calibration results

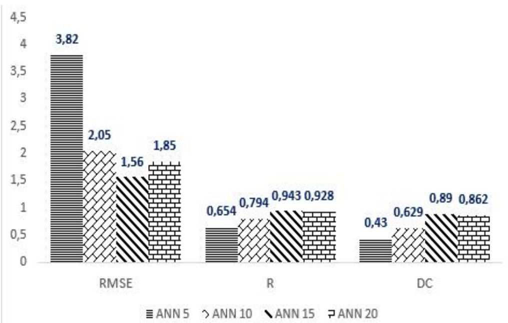

Fig. 6. Models Validation results

Specifically, we varied the input data for our ANN from 5 to 20 potability criteria in steps of 5, i.e. 5, 10, 15 and 20 criteria.

Thus, the ANN20 model with 20 input criteria gives us RMSE, R and DC calibration values of 1.71, 0.815 and 0.753 respectively, and validation values of 1.85, 0.928 and 0.862. The absolute difference between calibration and validation, for RMSE, R and DC, gives us 0.14, 0.113 and 0.109

respectively. For the ANN15 model, the RMSE, R and DC calibrations give values of 1.57, 0.945 and 0.910 respectively, and the validation values are 1.56, 0.943 and 0.890 respectively (Fig. 7). The absolute difference between calibration and validation for RMSE, R and DC is 0.01, 0.02 and 0.02 respectively. With the ANN10 model we have respectively 1.89, 0.84 and 0.687 in calibration and 2.05, 0.794 and 0.629 in validation (Fig. 8). For the ANN5, the calibration gives us 2.35, 0.793 and 0.587 and 3.82, 0.654 and 0.43 in validation (Fig. 9).

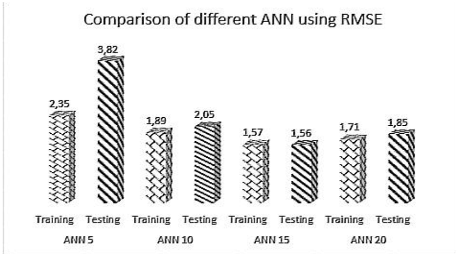

Fig. 7. Validation of RMSE Model

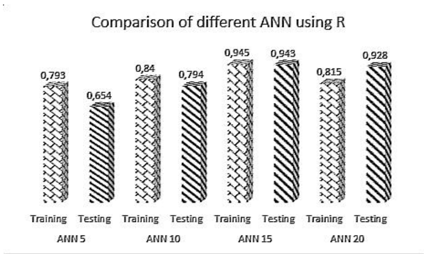

Fig. 8. Validation of R Model

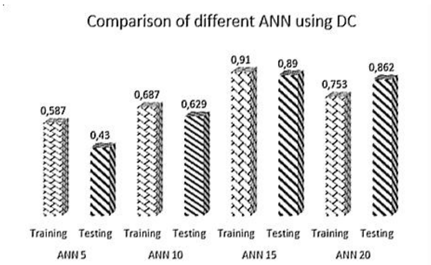

Fig. 9. Validation of DC Model

This first phase of our analysis allows us to affirm that the RMSE produces the best prediction with deviations between 0.01 and 0.02.

The second analysis of the variation in the input criteria using the RMSE predictive model shows an absolute difference between calibration and validation of 1.47, 0.16, 0.01 and 0.14 for the inputs ANN5, ANN10, ANN15 and ANN20 respectively. This comparative study of absolute deviations gives an overall variation of between 1.47 and 0.01, with a minimum threshold of 0.01 for the ANN15 model (Fig 7).

Algorithm 1 Variables Selection

Inputs : N = x1, ..., xm /* The itemset of potential variables
r $\leq$ m /* Number of selected variables
W /* Suitability measure
Outputs : S, C, N, i /* All selected variables
S $\leftarrow$ R
for $i = 1$ **to** $(m - r)$ **do**
  Xn $\leftarrow$ W(x_n)
  S $\leftarrow \frac{S}{x_n}$
  N $\leftarrow \frac{N}{x_n}$
end for

Algorithm 2 Perceptron_Training

function PERCEPTRONTRAINING(examples, network) $\triangleright$ **returns** perceptron hypothesis
  Inputs : examples a set of examples, X = $x_1,...,x_n$
  Outputs : **y** , **network**, perceptron weights W_i, **j = 0...n**, and activation function **g**
  repeat
    for each e **in** $examples$ **do**
      in $\leftarrow \sum_{n}^{j=0} W_j x_j[e]$
      Err $\leftarrow y[e] - g(in)$
      $W_j \leftarrow W_j + \alpha \times g(in) \times x_j[e]$
    end for
  until some stopping criteria is satisfield
    return PerceptronHypothesis(**network**)
  end function

Vol. 4, No. 2, July 2018, pp. xx-xx

Algorithm 3 Water_Prediction *using the neural network*

nb, nbmax /* All selected variables
Set the maximum number of repeats (nbmax) /* Each selected number of neurons
Outputs :
for $nb = 1$ **to** nbmax **do**
  repeat
    Select initial weight values (W_i)
    Present the learning base to the ANN (Lb_i)
    if $W_i \cong Lb_i$ **then**
      Calculate on sample
      Eq. 4
      Eq. 3
      Eq. 5
    else
      return to 1
    end if
    Compare the result and get the best for this model
  until nb $\leq$ nbmax
    Select the best of each selected model
  end for

Algorithm 4 WaterPrediction

/* Generate set of criteria and their parameters
Call Algorithms 1 2, 3, 3
/* Quality assessment of candidate criteria
function QUALITYEVALUATION(PEC,i,C)
  $Q_{PEC_m} \leftarrow \frac{\sum_{i=1}^{N} \frac{C_i}{PEC_i}}{N}$
  return Q_{PEC_m}
end function
function WATERQUALITY(N, Q, w, v, S)
  $w_i \leftarrow \frac{1}{\sum_{i=1}^{N} S_i}$
  $q_i \leftarrow 100 \times (\frac{V_i - V_{fine}}{S_i - V_{fine}})$
  $WQI \leftarrow \frac{\sum_{i=1}^{N} q_i \times w_i}{\sum_{i=1}^{N} w_i}$
  return WQI
end function

4. Conclusion

The main objective of this study was to analyse the ability of multilayer perceptron ANNs to inhibit pollution of an artesian source in contact with runoff water. Specifically, it should identify the best approaches for determining the criteria and parameters for predicting the potability of the said source. This prediction focused on 63 potability criteria grouped into 5 parameters.We compared 3 ensemble learning models using Big data in a comprehensive way. The correlation between key parameters and water quality was also identified and validated. The main conclusions of this study are as follows :

• Big data from wastewater treatment plants can be used to improve the performance of learning models in predicting the water quality of artesian sources;

• The correlation between the content of pollutants and the amount of inhibitors to annihilate the effect of waterborne pollutants shows a significantly better prediction performance;

• Two sets of key water parameters were identified and validated by the learning models.

In summary, the learning model of key water treatment parameters identified and validated by Big data in this study should be recommended for future monitoring of artesian source quality. They will contribute to water quality prediction, reduce the cost of artesian source and groundwater treatment and provide alerts.

We have achieved our objective of demonstrating the performance of ANNs in machine learning. We can conclude that ANN15 gives better results on the problem with deviation indicators of 0.01 for RMSE, 0.02 for R and DC between the calibration and the validation. However, the result is still not satisfactory due to the size of the data used. To improve this result, it would be interesting to introduce other variables that will affect the output to be predicted.

5. References

[1] Lasserre, F. (2021). Les enjeux politiques de l'eau: au-dela du cliché des conflits, la question de la gouvernance. Les Territoires de l'eau.

[2] Maeseele, C., Pradinaud, C., Payen, S., & Roux, P. (2021). L'empreinte eau-Mémento graphique.

[3] https://www.who.int/fr/news-room/fact-sheets/detail/drinking-water 24.08.2022 12:55

[4] Lowe, M., Qin, R., & Mao, X. (2022). A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring. Water, 14(9), 1384.

[5] Malik, A., & Bhagwat, A. (2021). Modelling groundwater level fluctuations in urban areas using artificial neural network. Groundwater for Sustainable Development, 12, 100484.

[6] Organisation Mondiale de la Santé, OMS. (2017). Directives de qualité pour l'eau de boisson : 4e ed. intégrant le premier additif.

[7] DECRET N°2019/ -0204/PRES/PM/MEA/MINEFID/MATDC/MS portant définition des normes, critéres et indicateurs d'accès à l'eau potable.

[8] https ://fr.unesco.org/themes/eaux-souterraines 24.08.2022 12:55

[9] https ://oneabf.com/tout-sur-leau/ 28.08.2022 13:50

[10] Kriech, A. J., & Osborn, L. V. (2022). Review of the impact of stormwater and leaching from pavements on the environment. Journal of Environmental Management, 319, 115687

[11] Horton, R. K. (1965). An index number system for rating water quality. Journal of Water Pollution Control Federation, 37(3), pp 300-306

[12] Brown, R. M., McClelland, N. I., Deininger, R. A., Tozer, R. G. (1970). A Water Quality Index- Do We Dare ? Water and Sewage Works, 117, pp. 339–343

[13] Khan, A., & Qureshi, F. R. (2018). Groundwater quality assessment through water quality index (WQI) in New Karachi Town, Karachi, Pakistan. Asian Journal of Water, Environment and Pollution, 15(1), 41-46.

[14] Philip, C., & Chhabra, V. (2022). Arsenic contamination : Sources, Extent and Impact on the Environment. EM International. Poll Res. 41 (1) : 312-322. DOI No.: http://doi.org/10.53550/PR.2022.v41i01.046

[15] Huq, M. E., Fahad, S., Shao, Z., Sarven, M. S., Khan, I. A., Alam, M., ... & Khan, W. U. (2020). Arsenic in a groundwater environment in Bangladesh: Occurrence and mobilization. Journal of environmental management, 262, 110318.

[16] Telesphore Tiendrebeogo, Mamadou Diarra, (2020) "Big Data Storage System Based on a Distributed Hash Tables system" International Journal of Database Management Systems (IJDMS) Vol.12, No.4/5.

[17] Sirisha Adamala. An Overview of Big Data Applications in Water Resources Engineering. Machine Learning Research. Vol. 2, No. 1, 2017, pp. 10-18. doi : 10.11648/j.mlr.20170201.12

[18] Ponce Romero, J. M., Hallett, S. H., & Jude, S. (2017). Leveraging big data tools and technologies : addressing the challenges of the water quality sector. Sustainability,9(12), 2160.

[19] Enemark, T., Peeters, L. J., Mallants, D., & Batelaan, O. (2019). Hydrogeological conceptual model building and testing : A review. Journal of Hydrology, 569, 310-329.