



# An optimized explainable artificial intelligence approach for sustainable clean water

Dalia Ezzat<sup>1,2,3</sup> · Mona Soliman<sup>1,3</sup> · Eman Ahmed<sup>1,3</sup> · Aboul Ella Hassanien<sup>1,3</sup>

Received: 9 February 2023 / Accepted: 27 July 2023  
© The Author(s) 2023

## Abstract

Water, sanitation, and hygiene are essential components of the 2030 agenda for sustainable development. Goal 6 is dedicated to guarantee all societies have access to water and sanitation. Water quality (WQ) assessment is crucial to ensure the availability of clean water. This paper presents an approach called AHA–XDNN for predicting WQ. The proposed approach is based on three pillars to predict WQ with high accuracy and confidence, namely, deep neural networks (DNN), artificial hummingbird algorithm (AHA), and explainable artificial intelligence. The proposed approach involves five phases: data preprocessing, optimization, training, and evaluation. In the first phase, problems such as unwanted noise and imbalance are addressed. In the second phase, AHA is applied to optimize the DNN model's hyper-parameters. In the third phase, the DNN model is trained on the dataset processed in the first phase. The performance of the optimized DNN model is evaluated using four measurements, and the results are explained and interpreted using SHapley additive exPlanations. The proposed approach achieved an accuracy, average precision, average recall, average F1-score of 91%, 91%, 91.5%, and 91% on the test set, respectively. By comparing the proposed approach with existing models based on artificial neural network (ANN), the proposed approach was able to outperform its counterparts in terms of average recall and average F1-score.

**Keywords** Explainable artificial intelligence (XAI) · Water quality · Imbalanced data · Artificial hummingbird algorithm (AHA) · SHapley additive exPlanations (SHAP)

---

✉ Dalia Ezzat  
dalia\_ali@cic-cairo.com

Mona Soliman  
mona.solyman@fci-cu.edu.eg

Eman Ahmed  
e.ahmed@fci-cu.edu.eg

Aboul Ella Hassanien  
aboitcairo@cu.edu.eg

<sup>1</sup> Faculty of Computers and Artificial Intelligence, Cairo University, Cairo, Egypt

<sup>2</sup> School of Computer Science, Canadian International College (CIC), Cairo, Egypt

<sup>3</sup> Scientific Research Group in Egypt (SRGE), Cario, Egypt

## 1 Introduction

In response to human ambitions for a better life and the potential repercussions on the planet's sustainability, the United Nations developed the 2030 agenda for sustainable development, which encompasses 17 sustainable development goals (SDGs) and 169 targets. Among the 17 goals, goal 6 is dedicated to the provision and management of water and sanitation for all, highlighting the importance of water on the global political agenda. Because water is essential to human health and life, as well as the welfare and sustainability of the planet, it is also a main condition for development. Water is essential for coping with climate change, as it serves as a link between the environment, the climate system, and human society (Delanka-Pedige et al., 2021; Nhamo et al., 2019). But as urbanization and industrialization expand, a number of harmful chemicals are produced (Jahangard et al., 2022). Heavy metals such as mercury, organic pollutants such as pesticides and microbiological illnesses are just a few of the forms in which the manufactured wastes are released into water bodies.

All of these water contaminants are harmful to local wildlife and people's health. There is an urgent need to monitor WQ levels. Water is said to be of good quality if it is free of potentially dangerous biological forms and organisms that might be distasteful. It is translucent, colorless, and has no flavors or odor. It does not include any chemical concentrations that might be harmful to your health, unsightly, or disastrous to your finances. Each year, many people suffer from kidney failure, cancer, etc., as a result of contaminated water (Abdulla, 2021). Classification of WQ in laboratories requires laborious processes and a lot of resources. Currently, there are many ways for classifying WQ; however, they are not accurate. The majority of research uses two sorts of studies to help determine WQ: regular laboratory tests and data analysis. Having an automated system that can quickly and easily assess the WQ is therefore imperative.

Recently, artificial intelligence (AI) has provided automated methods for problem-solving using vast amounts of data that cannot be processed by humans for the purpose of decision-making, including equipment selection, operation optimization, and problem-solving. AI techniques can effectively duplicate this behavior and make up the deficit for this purpose. Different studies use AI-based methodologies to assist identify the most efficient approach to solve the WQ problem. The challenge of predicting WQ has started to be solved using traditional machine learning (ML) approaches (Khoi et al., 2022). ML is a branch of AI which refers to a system's capacity to gather, combine, and create knowledge from massive amounts of data without programming. Support vector machine (SVM), decision tree (DT), random forest (RF), and adaptive boosting (AdaBoost) are examples of ML models.

As a response to the limitations of ML, deep learning (DL) was developed. DL networks such as ANN, recurrent neural network (RNN), and convolutional neural network (CNN) can tackle the industrial issue of executing sundry operations on big amounts of data. Despite promising results in predicting WQ in either DL- or ML-based approaches, the lack of transparency in the currently presented AI-based methods, which prevents the evaluation of model outputs, is a fundamental defect in the validity and fairness of these models. Explainable AI (XAI), which describes an AI model, its expected impact, and any prospective biases, has lately been utilized to address this issue. It contributes to determining the validity, reliability, and transparency of AI models. An organization must first build trust and confidence before using AI models in

production. An organization can adopt a responsible and ethical approach developing traditional AI models with the assistance of XAI.

Therefore, this paper introduces XAI approach-based deep neural network (DNN) and artificial hummingbird algorithm (AHA) for predicting WQ. The proposed XAI approach consists of five phases, namely data pre-processing phase, optimization phase, training phase, model evaluation phase, and results explanation phase. In the data pre-processing phase, the dataset used is processed from undesirable noise and imbalance. In binary classification problems, data imbalance refers to the case when one class has more samples than the other class resulting in a dominated class and a minority class. This causes the classifiers to give biased results in favor of the majority class. Several methods exist for handling this problem including random under-sampling and random oversampling (Johnson & Khoshgoftaar, 2019). In the under-sampling method, randomly chosen samples from the dominant class are eliminated. This results in decreasing the number of the majority class to reach that of the minority one. In the oversampling method, a randomly samples from the minority class are repeated so that the number of samples becomes equal to the number of samples in the dominant class. The problem with random oversampling is that the repeated samples do not append any additional knowledge to the model. As an alternative, the synthetic minority oversampling technique (SMOTE) oversampling (Chawla et al., 2002) synthesizes similar samples to those of the minority class. In the optimization phase, the AHA is employed to choose the DNN's hyper-parameters' ideal values that significantly affect its results before using it in the next phase in WQ prediction. In the training phase, the DNN model learns from the dataset that was processed in the first phase and then presented and then its results are presented and interpreted in the model evaluation and results explanation phases, respectively. A list of this paper's significant contributions is provided below:

- An approach based on AHA and explainable deep neural network (XDNN) is presented to address the problem of WQ prediction.
- AHA is utilized to optimize the hyper-parameters of XDNN to increase prediction performance.
- The results of the proposed approach AHA–XDNN are very competitive, achieving an accuracy level of 91% in the test set.
- The XAI technique SHAP was adopted to explain the internal prediction mechanism of the AHA–XDNN approach.

The breakdown of the paper's remaining sections is as next. Section 2 offers a review of numerous cutting-edge models for WQ prediction including AI-based models and XAI-based models. Section 3 presents the relevant theories and detailed information about the dataset that was utilized in this paper. Section 4 presents the proposed approach, Sect. 5 offers the experimental findings, the conclusion and discussions about future work can be found in Sect. 6.

## 2 Literature review

There is a mounting interest in assessing WQ; there has been a larger need for trustworthy, precise, and adaptive prediction models. This paper's primary goal is to predict WQ in order to ensure that it is safe to consume. In order to meet the need for adaptable models,

in this paper an approach that does not require human intervention is provided, in which an optimization algorithm was utilized to choose the ideal values for the hyper-parameters of the DL model used. Moreover, an XAI method has been used to guarantee the reliability and validity of the proposed approach's results. Lately, a number of researchers have employed AI systems to predict WQ, with encouraging results. Some of these systems have incorporated XAI methods to interpret the obtained results. However, at the time of writing this paper, an approach similar to the one presented in this paper that is based on hyper-parameters optimization of DL models with incorporation of interpretation methods has not been proposed. In this section, recently developed models that employ either AI-based methods or XAI-based methods will be explored.

## 2.1 AI-based methods for water quality prediction

Different ML techniques have been utilized to address WQ prediction problems. Khan and See (2016) has created a WQ prediction model utilizing time series analysis and ANNs. The WQ was estimated by Khan and See (2016) utilizing 12 ML models. As a way to assess the performance of each model, different methods of regression analysis were used such as root-mean-squared error (RMSE).

Yahya et al. (2019) attempts to develop a useful model using SVM to assess the WQ by analyzing six parameter data for twin reservoirs situated in the watershed. The primary gain of the suggested model is that catchments without gauges or with insufficient monitoring stations for WQ indicators may find it useful. Nair and Vijaya (2021) used a variety of ML and big data approaches utilizing sensor network-based prediction models. Several methods have been applied in Hassan et al. (2021) to forecast WQ such as RF, XGB, DT, and Ada-Boost, and among the models used, XGB yielded the highest accuracy of 83%. The challenge of having outliers in the dataset was resolved, and the accuracy of the WQ prediction was increased by the automatic WQ prediction method in Juna et al. (2022). Juna et al., (2022) recommended resolving the missing value issue by combining a multilayer perceptron (MLP) of a nine-layer and a K-nearest neighbor (KNN) imputer. Panigrahi et al., (2023) proposed a ML-based model for the purpose of predicting ground WQ for drinking suitable in accordance with WHO guidelines; the intended issue is expressed as a multiclass categorization issue. AI approaches like decision trees, Ada-Boost, KNN, XGBoost, logistic regression, and many SVM versions were used. Results reported in Panigrahi et al. (2023) showed that Ada-boost, XGBoost, and the polynomial SVM model all correctly identified the WQ classes, according to prediction findings. It would assist in selecting the safest source of drinking water.

Many DL models have lately been employed to address the problem of WQ prediction. An ANN-based model was proposed in Rustam et al. (2022) for predicting WQ and water consumption. The ANN model was able to achieve very accurate and reliable results for estimates of WQ and water use. It yielded an accuracy of 0.96 for predicting WQ and achieved 0.997 R2 for producing water consumption, outperforming other approaches with these results. On the other hand, CNN is unable to learn sequence association. Long short-term memory (LSTM) architecture is particularly evolved to handle issues that are closely related to time series because of its superior information memory and sequential modeling capabilities, such as process monitoring. Various WQ prediction models based on LSTM and DNN were introduced in Wang et al. (2017), Bi et al., (2021), Farhi et al., (2021), Venkata Vara Prasad et al., (2022), Wang et al., (2023), Zhao et al., (2020), Zheng et al., (2021), Qin et al., (2017), Liang et al., (2018); Rasheed

Abdul Haq and Harigovindan (2022), and Zhou et al., (2018). The academics presented a unique feature selection and categorization system in Charles et al. (2021) for precise real-time WQ prediction. The complexity of the proposed system is reduced by selecting the optimal set of attributes utilizing a learning-based model and quantum teaching.

## 2.2 XAI-based methods for WQ prediction

The XAI is a cutting-edge technique that offers an explanation of a ML model's outcomes based on its features the connections between these features. Consequently, it gets around a key problem with black box-based ML models and makes advanced ML models more useful (Adadi & Berrada, 2018). In Park et al., (2022a) investigation, SHAP, a unique XAI technique, was employed to evaluate the model's results and offer a clear explanation of the predictions. In this work, the performance of the model's input variables was interpreted in an understandable manner using SHAP analysis. The SHAP values in the XGB model reflect the affined weighting of the input features (Park et al., 2022a). An XGB model was created to forecast the pace at which the WQ in a water treatment facility recovered following a disruption to the water treatment process. A pre-processing steps were conducted on the used data to enhance the model's prediction based on how the recovery rate was defined. Additionally, a brand-new XAI technique has been applied to examine the model's findings. An acceptable interpretation of the model's outcomes was supplied by the study of model predictions utilizing the SHAP values for and the target plots of the input features. The findings of this study show how a ML model may be utilized to predict recovery in water treatment operations following errors. In light of the features of the input variables, the significance of pre-processing of the data utilized in the model building has also been underlined. This study's suggested approach offers a helpful strategy for more reliable and efficacious control of water treatment systems. A comparison of various ML techniques, including SVM, DT, Ada-Boost, and RF, utilized for the classification of WQ is presented in Patel et al. (2022). The WQ index dataset from Kaggle is utilized to train each model. The dataset is normalized and balanced using Z-score and SMOTE, respectively, before the model is started to be trained. The findings of the experiments indicate that RF and gradient boost provide a maximum accuracy of 81%. In order to decide which aspects are most crucial, the authors employed XAI. To determine the effect of each feature in the obtained findings, local interpretable model-agnostic explanations (LIME) are used.

This work (Park et al., 2022b) effectively illustrated a solid example of how to apply XAI to enhance the explication of ML model's results in forecasting WQ. The influence of input feature selection on model's output was assessed, with the three indicators SHAP, feature importance (FI), and variance inflation factor (VIF) being used to rank the relevance of input variable selection. The study demonstrates that the model's performance is consistently better when relying on SHAP to determine the order of importance of the input variables. This reveals that it is possible to lower the cost of the entire WQ analysis by designing on-site monitoring to gather the chosen input variables from the SHAP analysis. The study in Madni et al. (2023) used also the SHAP to explain the significance of various features after applying the stacked ensemble H<sub>2</sub>O AutoML model and utilizing the KNN imputer to handle the omitted values. Several learning models were used in experiments to analyze the effectiveness of the KNN imputer and the suggested H<sub>2</sub>O AutoML model.

### 3 Material and methods

#### 3.1 Artificial hummingbird algorithm

The AHA is a metaheuristic method for handling optimization problems and was developed in Zhao et al. (2022). The initialization and foraging phases make up the two main stages of the AHA algorithm. There are three foraging strategies which are the guided, territorial, and migration foraging. Next, an explanation will be provided for each of the stages. In the Initialization phase by using the equation below, a population of  $N$  randomly initialized hummingbirds is set on  $N$  food places:

$$z_i = L + r \cdot (U - L) \quad i = 1, \dots, N \quad (1)$$

where  $z_i$  denotes the location of the  $i$ th food place that is the solution of a given problem,  $L$  and  $U$  are the lower and upper bounds for a  $d$ -dimensional problem, respectively, and  $r$  is a random vector in the range  $[0, 1]$ . Each hummingbird favors the food place with the greatest visit level when choosing where to forage. The initialization of this visit table is as in the following way:

$$VT_{i,j} = \begin{cases} 0 & , \text{if } i \neq j \\ \text{null} & , \text{if } i = j \end{cases}, \quad i = 1, \dots, N \text{ and } j = 1, \dots, N \quad (2)$$

where for  $i = j$ ,  $VT_{i,j} = \text{null}$  denotes that a hummingbird takes food at its particular food place; for  $i \neq j$ ,  $VT_{i,j} = 0$  means that the  $j$ th food place has just been explored by the  $i$ th hummingbird in the current iteration.

As mentioned above, there are primarily three foraging strategies: guided, territorial, and migration. In guided foraging, the target food place is one that the hummingbird has not visited in a while because it is already full and has the highest rate of refilling nectar. After consuming food from the intended food place, territorial foraging begins, during which hummingbirds seek out new food places rather than going to other known food places. In an attempt to find a neoteric food place that is richer than the current one, it makes an effort to travel to a nearby location. The hummingbird has 50% probability of making a guided or territorial foraging. When the most frequently frequented area is deficient in food, migration foraging takes place. The hummingbird will then depart its zone and search for a more remote food place. Figure 1 illustrates the three foraging strategies.

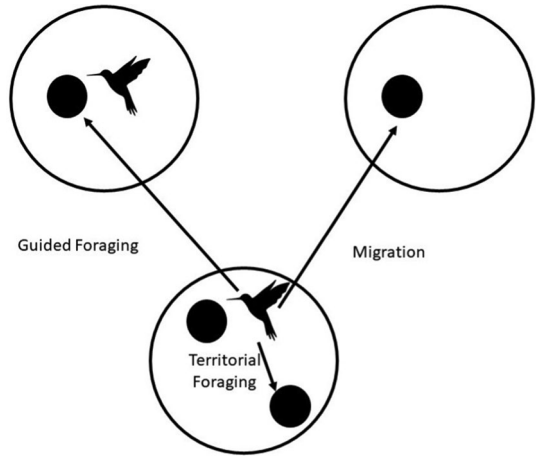
The hummingbird has three different flight patterns when it is foraging. These flights are axial, diagonal, and omnidirectional. A hummingbird can fly along any coordinate axis. Any flying direction could be projected to each of the three coordinate axes, as seen by the omnidirectional flight. In a  $d$ -dimension space, the axial flight is computed as Eq. (3), the diagonal flight is represented as Eq. (4), and the omnidirectional flight is given as Eq. (5).

$$\varphi_i = \begin{cases} 1 & \text{if } i = \text{randi}([1, d]) \\ 0 & \text{else} \end{cases}, \quad i = 1, \dots, d \quad (3)$$

$$\varphi_i = \begin{cases} 1 & \text{if } i = P(j), j \in [1, k] \\ 0 & \text{else} \end{cases}, \quad P = \text{rand perm}(k), k \in [2, r_1 \cdot (d - 2) + 1], i = 1, \dots, d \quad (4)$$

$$\varphi_i = 1, i = 1, \dots, d \quad (5)$$

**Fig. 1** The foraging strategies. The food place is depicted by the black circle



where  $\text{randi}([1, d])$  produces a random integer from 1 to  $d$ ,  $\text{randperm}(k)$  gets a random permutation of integers from 1 to  $k$ , and  $r_1$  is a random number in range  $[0, 1]$ . The flight pattern of a hummingbird affects the choice of candidate food places. In guided foraging, the following mathematical equation is to determine a candidate food place  $v_i(t + 1)$ :

$$v_i(t + 1) = z_{i,\text{tar}}(t) + a \cdot \varphi \cdot (z_i(t) - z_{i,\text{tar}}(t)) \tag{6}$$

$$a \sim N(0, 1) \tag{7}$$

where  $z_{i,\text{tar}}(t)$  is the location of the intended food place that the  $i$ th hummingbird likes to visit.  $z_i(t)$  is the location of the  $i$ th food place at time  $t$ ,  $\varphi$  is the flight pattern vector, and  $a$  is a guided factor that follows the normal distribution  $N(0, 1)$  with mean = 0 and standard deviation = 1. In territorial foraging, a food place  $v_i(t + 1)$  is discovered as:

$$v_i(t + 1) = z_{i,\text{tar}}(t) + b \cdot \varphi \cdot z_i(t) \tag{8}$$

$$b \sim N(0, 1) \tag{9}$$

The location update rule of the  $i$ th food place is given as:

$$z_i(t + 1) = \begin{cases} z_i(t) & f(z_i(t)) \leq f(v_i(t + 1)) \\ v_i(t + 1) & f(z_i(t)) > f(v_i(t + 1)) \end{cases} \tag{10}$$

where  $z_i(t)$  refers to the  $i$ th food place at iteration  $t$  and  $f$  denotes the function fitness value. The migration foraging occurs when a hummingbird departs to a more distant food place since the area it often visits is likely to be food scarce. The hummingbird at the food place with the lowest rate of nectar refilling will move to a neoteric food place generated at random throughout the search area when the amount of iterations exceeds the stated value of the migration coefficient. In relation to population size, the migration coefficient is calculated as follows:

$$M = 2N \tag{11}$$

The visit table is then adjusted as the hummingbird switches from the old source to the new one at this moment. One way to describe a hummingbird's journey foraging from one nectar source to another created randomly is as follows:

$$z_{\text{worst}}(t+1) = L + r \cdot (U - L) \quad (12)$$

where  $z_{\text{worst}}(t+1)$  stands for the new position after being in the place with the worst fitness value,  $L$  and  $U$  are the lower and upper bounds for a  $d$ -dimensional problem, respectively, and  $r$  is the randomization coefficient used to choose the new position. Figure 2 summarizes the AHA algorithm. It starts with the initialization phase and then calculation of the fitness value of the initial candidate solution. Then, the flight pattern is chosen followed by the foraging strategies. Iterations are performed until stopping criteria are met. In the end, the best reached solution is returned.

### 3.2 Explainable artificial intelligence (XAI)

An AI model is viewed as a “black box” that is capable of providing “yes” or “no” responses without elaborating on how they were arrived at. To guarantee trust and transparency, many applications require a justification of how an answer was generated. In order to make the black-box AI systems understandable, this gave rise to a new branch of AI research known as XAI (Gohel et al., 2021). The main objective is to deliver “wh” answers regarding an output. For instance, XAI should be able to respond to questions like “why a specific output was attained?” “how a specific output was attained?”, and “when a specific AI-based system can fail?” (Garcia et al., 2018; Neerincx et al., 2018; Zhou et al., 2018). Accordingly, the objectives of XAI are to provide transparency and trustworthiness to the used AI techniques in various applications. Transparent AI models must be expressive enough for humans to understand them. Trust can be acquired by having a logical and rational justification for any decision made by the AI model. The most famous XAI

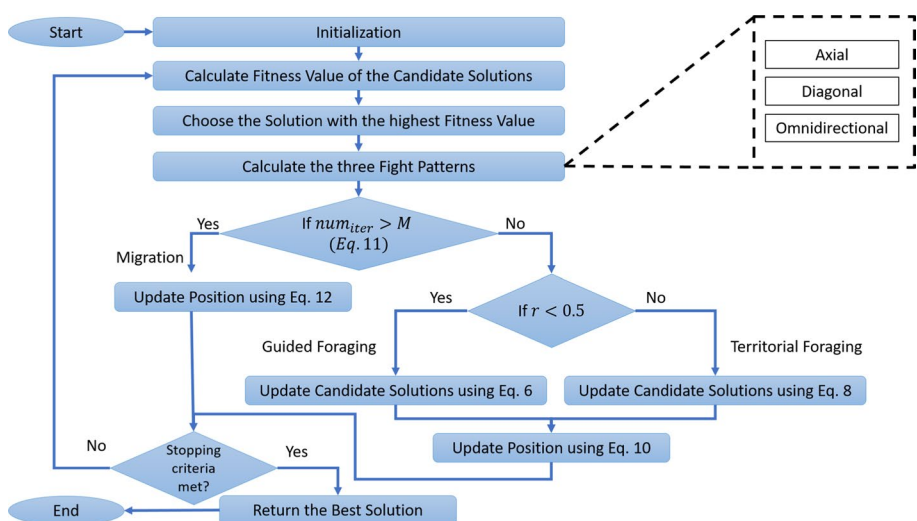


Fig. 2 Flowchart of AHA



techniques are the local interpretable LIME (Ribeiro et al., 2016) and SHAP (Lundberg & Lee, 2017).

The SHAP framework is a game theoretic approach for explainable AI. It enables us to explain the reason behind the outputs of ML models using Shapely values. SHAP utilizes coalitional game theory to determine Shapely values. It is a technique for rewarding game players in accordance with their contribution to the game, as described by Shapley (1953). In AI, the input features act as the players and the model’s decision is considered as the game outcome. Applying SHAP explains how every feature in the input data contributes to every prediction. Given a group of features  $F$ , a set of possible features coalitions  $S \subseteq F$  results in output  $O(S)$ . The Shapely value ( $\varphi_i$ ) is the average marginal contribution of a specific feature  $i$  acquired using various possible coalitions; it is given by:

$$\varphi_{\text{cont}}(i) = \sum_{S \subseteq F \setminus \{i\}} c \cdot \text{cont}(i, S) \tag{13}$$

$$C = \frac{|S|!(p - |S| - 1)!}{p!} \tag{14}$$

$$\text{cont}(i, S) = (\text{cont}(S \cup \{i\}) - \text{cont}(S)) \tag{15}$$

$C$  is a normalization term that considers the number of choices for the subset  $S$ ;  $\text{cont}(i, S)$  is feature  $i$ ’s marginal contribution w.r.t. coalition  $S$ .

### 3.3 Dataset characteristics

The WQ dataset that employed in this paper was acquired from Kaggle (<https://www.kaggle.com/datasets/mssmartypants/water-quality>). The dataset consisted of 21 features and 7999 samples. Table 1 depicts the features of the WQ dataset, and the ranges and unsafe limits of each feature.

## 4 Proposed approach

The proposed AHA–XDNN approach consists of five phases as depicted in Fig. 3. Data pre-processing is the first phase, followed by the optimization phase, the training phase, the model evaluation phase, and the results explanation phase. The proposed approach for predicting WQ is based on DNN. DNN comprised in several hidden layers between the input layer and the output layer (Awad & Khanna, 2015). For classification, the number of neurons in the output layers is equivalent to the number of classes. Through synapses, each layer’s neurons are connected to the subsequent layer’s neurons. Each synapse has a weight for neurons activation. During training, the network learns the values of the weights to learn a certain function. The network learns from training samples, while how much it has learned is tested using test samples. The training samples contain data samples and their corresponding class labels. The test samples are usually samples with unknown class labels. In addition, a DNN can contain many different types of layers than previously mentioned. These layers can be such as the activation layer that proposes nonlinearity, drop-out layers used for regularization (Srivastava et al., 2014), batch normalization (Ioffe & Szegedy, 2015) utilized to normalize the outputs of the neurons, and was found to have a

**Table 1** Features of the WQ dataset and its ranges and impermissible limits

Features	Range per liter of water	Impermissible limits
Aluminum	[0, 5.05]	Above 2.8
Ammonia	[-0.08, 29.8]	Above 32.5
Arsenic	[0, 1.05]	Above 0.01
Barium	[0, 4.94]	Above 2
Cadmium	[0, 0.13]	Above 0.005
Chloramine	[0, 8.68]	Above 4
Chromium	[0, 0.9]	Above 0.1
Copper	[0, 2]	Above 1.3
Fluoride	[0, 1.5]	Above 1.5
Bacteria	[0, 1]	Above 0
Viruses	[0, 1]	Above 0
Lead	[0, 0.2]	Above 0.015
Nitrates	[0, 19.83]	Above 10
Nitrites	[0, 2.65]	Above 1
Mercury	[0, 0.01]	Above 0.002
Perchlorate	[0, 60.01]	Above 56
Radium	[0, 7.99]	Above 5
Selenium	[0, 0.1]	Above 0.5
Silver	[0, 0.5]	Above 0.1
Uranium	[0, 0.09]	Above 0.3
Is_safe	Target class (0=not safe, 1= safe)	

good impact on the model accuracy. Although the use of these additional layers improves the results of DNN models, the values of the hyper-parameters of any DNN model significantly affect the performance and results. If inappropriate values are set for these hyper-parameters, this will negatively affect the results (Darwish et al., 2020). To overcome this hurdle, in the second phase of the proposed approach, the hyper-parameters of the used DNN model are tuned to their ideal values using AHA. In the following subsections, each phase will be thoroughly explained.

#### 4.1 Data preprocessing phase

In data preprocessing phase, first a statistical analysis of the WQ dataset was performed as shown in Table 2. Through this analysis unwanted noise was observed in the dataset as three samples in WQ dataset had missing values in the ammonia feature and target label. As the number of missing values is not many and ineffective, the three indicated samples were removed from the dataset. From the statistical analysis, it was noticed that the distribution of values in the dataset varies from 0.0 to 60.01. Therefore, the entire dataset was normalized to make all the values in it have the same distribution. To deal with the imbalance ratio of the dataset which is 7.76, 2000 samples were picked at random from the dominated class and the number of samples in the minority class was increased to 2000 samples using SMOTE. After making the dataset balanced, it was split into training set, validation set, and test set. The training set comprises 70% of the total samples of the

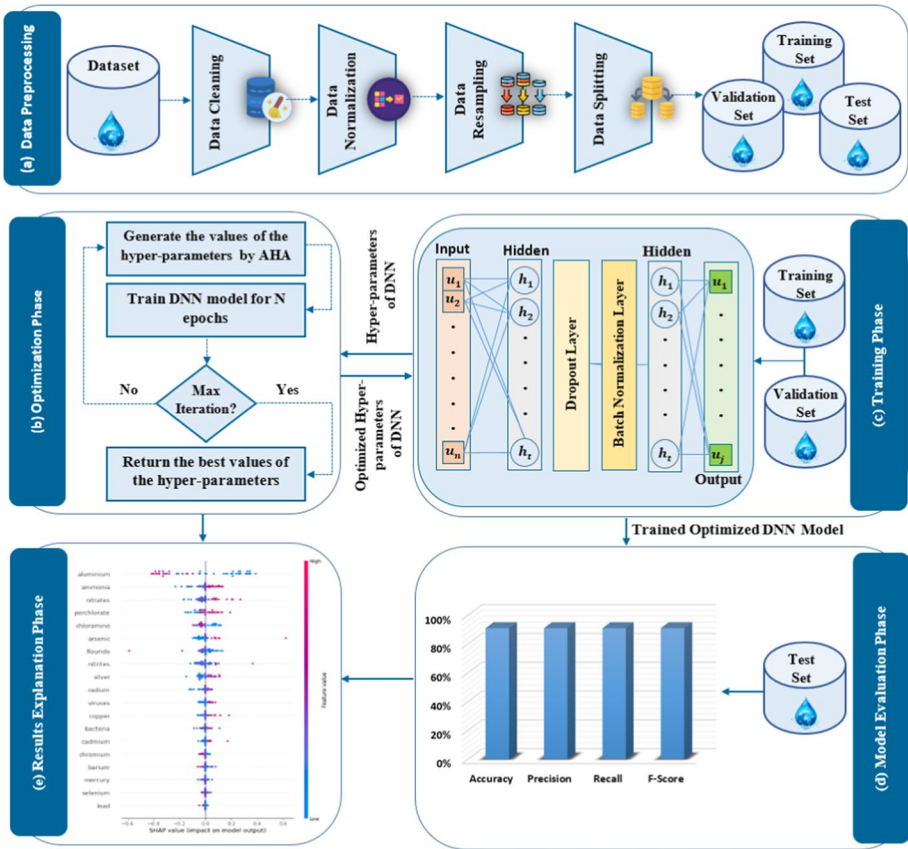


Fig. 3 The architectural form of the proposed approach

Table 2 Summary statistical analysis of the WQ dataset

WQ dataset	
Class distribution	0: 7084, 1: 912, undefined: 3
Imbalance ratio	7.76

dataset, the validation set includes 15% of the total samples of the dataset, and the test set includes 15% of the total samples of the dataset.

### 4.2 Optimization of the hyper-parameters phase

At the second phase of the proposed approach, the AHA is implemented to optimize the hyper-parameter values of the 6-layer DNN model used in the proposed approach. The DNN model consists of six layers as depicted in Fig. 3, which are the input layer, the output layer activated by the sigmoid function, and there are four layers between them. These four layers

are two hidden layers activated by ReLU function, one dropout layer and one batch normalization layer. Each of the hidden layers and the dropout layer is tied to a hyper-parameter. The first hidden layer is associated with a hyper-parameter called number of neurons  $FN_n$ ; the second hidden layer is also associated with the number of neurons  $SN_n$ , where the  $FN_n$  stands for the number of neurons in the first hidden layer and  $SN_n$  stands for the number of neurons in the second hidden layer of the DNN model. The dropout layer is tied to a hyper-parameter called the dropout rate  $D_r$ .  $FN_n$ ,  $SN_n$ , and  $D_r$  are all optimized using AHA. In other words, the search space is three-dimensional and each point in the space acts as a combination of these three hyper-parameters.

### 4.3 Training phase

At this phase, after the AHA determines the hyper-parameter values of the 6-layer DNN model, the 6-layer DNN model is trained for a number of iterations  $= N_i$  using the training set. The 6-layer DNN model's performance is assessed during the training on the validation set.

### 4.4 Model evaluation and results explanation phases

In the model evaluation phase, after the training of the 6-layer DNN model is ended, it is assessed on the test set. The 6-layer DNN model's performance is assessed utilizing a various of metrics including confusion matrix, accuracy, precision, recall, and F1-score. The most utilized criteria for gauging the effectiveness of classification models are accuracy. As indicated in Eq. (16), it is computed by counting the successfully categorized samples and then dividing this number by the total number of samples. As may be observed in Eq. (17), precision is defined by dividing the true positives by the total of true and false positives. As demonstrated in Eq. (18), recall, also known as sensitivity, is computed by dividing the true positives by the total of false negatives and true positives. F1-score is calculated using Eq. (19) and depends on precision and recall. It is employed to strike a balance between recall and precision. The confusion matrix can be thought of as a summary of a classifier's prediction outcomes. The confusion matrix sheds light on the classifier's mistakes and the kinds of mistakes made (Goutte & Gaussier, 2005; Tharwat, 2018; Ting, 2011).

$$\text{Accuracy} = \frac{T \text{ positive} + T \text{ negative}}{T \text{ positive} + T \text{ negative} + F \text{ positive} + F \text{ negative}} \times 100\% \quad (16)$$

$$\text{Precision} = \frac{T \text{ positive}}{T \text{ positive} + F \text{ positive}} \times 100\% \quad (17)$$

$$\text{Recall} = \frac{T \text{ positive}}{T \text{ positive} + F \text{ negative}} \times 100\% \quad (18)$$

$$F1 - \text{score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \quad (19)$$

where  $T$  positive = true positives,  $T$  negative = true negatives,  $F$  positive = false positives, and  $F$  negative = false negatives.

In the results explanation phase, the XAI method “SHAP” is used to explain the results of the 6-layer DNN model by figuring out how much each feature contributed to the prediction. The SHAP method has the ability to identify the most important features and their impact on model prediction.

## 5 Results and discussion

This part introduces and evaluates the proposed approach results. All operations of the proposed approach were implemented using Python and Keras (Chollet, 2015) and implemented on Google Colaboratory (Carneiro et al., 2018). The results section is divided into four sub-sections so that the results of data preprocessing phase, optimization phase, training phase, model evaluation phase, and results explanation phase of the proposed approach are clearly presented.

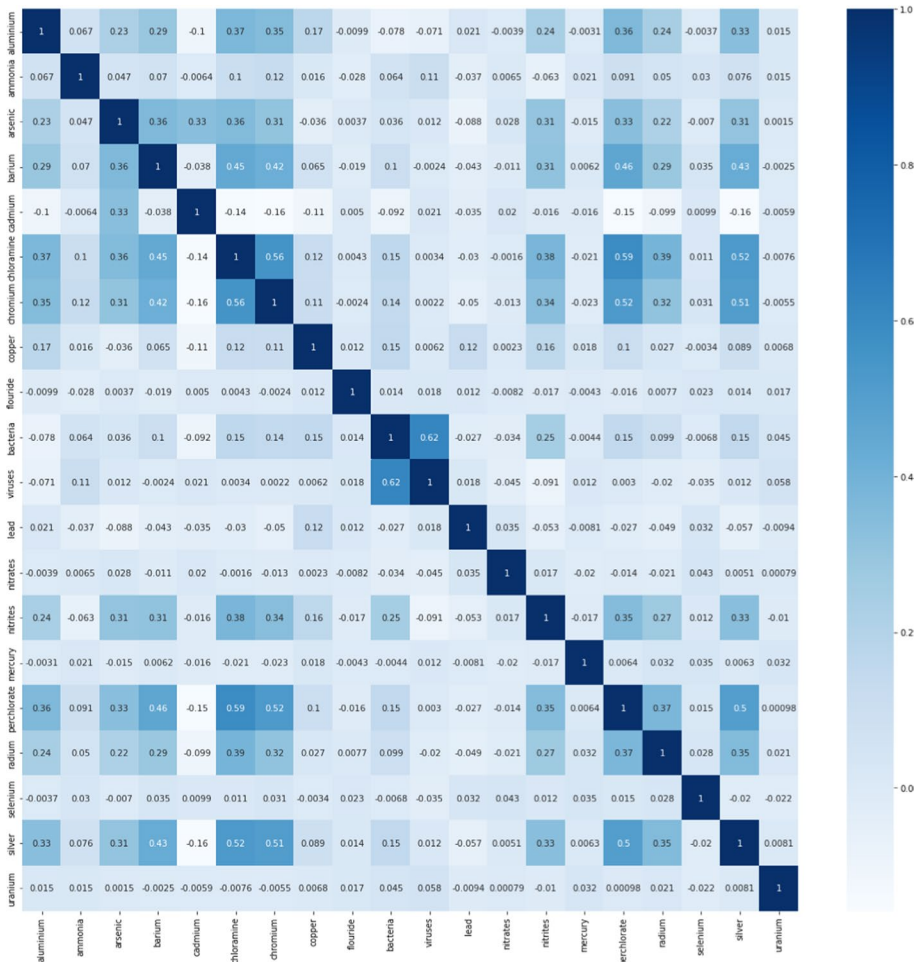
### 5.1 Data preprocessing phase of AHA–XDNN

In addition to the statistical analysis performed for the dataset, as described earlier in Sect. 4.1, it was also important to use the correlation matrix heatmap. The correlation matrix heatmap can provide a visual representation of which features of a dataset are most closely correlated to each other. Because highly correlated features add a degree of redundancy and also have an impact on the stability of any ML model. By examining the correlations in the heatmap of the WQ dataset as seen in Fig. 4, it was observed that the features of this dataset are not significantly correlated. The highest value of direct or positive correlation between the dataset’s features is only 0.62, and it is between “bacteria” and “viruses.” The highest value of inverse or negative correlation between the dataset’s features is only -0.16, and it is between “cadmium” and “silver.”

### 5.2 Optimization phase of AHA–XDNN

The search space for  $FN_n$ ,  $SN_n$ ,  $D_r$  whose values will be figured out by AHA has been constrained as follows. The search range for  $FN_n$  was restricted to [50, 1000], and the search range for  $SN_n$  was restricted to [50, 1000]. The search space for  $D_r$  was restricted to [0.1, 0.9], as shown in Table 3. The AHA hyper-parameters’ values were chosen at random, with population size and  $N_i$  tuned to 15 and 10, respectively. By experimenting with several values,  $N_i$  of 3-layer DNN model was determined. The results of the experiment revealed that the time of the optimization phase required significantly more time when more than ten iterations were employed. The results of the 3-layer DNN model were insufficiently precise when  $N_i$  of 3-layer DNN model was set to a value less than 10.

The AHA seeks to lower the three-layer DNN model’s loss rate on the validation set. In a more thorough explanation, after each iteration of the AHA, the appropriateness of the proposed solutions for the  $FN_n$ ,  $SN_n$ ,  $D_r$  is assessed based on the loss rate of the 3-layer DNN model on the validation set after training this model 10 iterations on the training set. When  $N_i$  of AHA reached 10, optimal values for  $FN_n$ ,  $SN_n$ , and  $D_r$  were determined. Table 4 displays the optimal values for the  $FN_n$ ,  $SN_n$ , and  $D_r$  determined by AHA, where the values for  $FN_n$ ,  $SN_n$ , and  $D_r$  were determined to be 400, 350, and 0.2, respectively.



**Fig. 4** Correlation matrix heatmap of the WQ dataset. The color value of the right bar from highest to lowest indicates that the features are not correlated

**Table 3** The values given to the hyper-parameters of the AHA and 3-layer DNN model during the optimization phase.  $N_i$ = number of iterations

Parameters	Assigned value
$N_i$ of AHA	10
Population size	15
$FN_n$	[50,1000]
$SN_n$	[50,1000]
$D_r$	[0.1, 0.9]
$N_i$ of 3-layer DNN model	10

**Table 4** Optimum values of the 3-layer DNN model’s hyper-parameters determined by AHA

Hyper-parameters of 3-layer DNN model	Optimal values
$FN_n$	400
$SN_n$	350
$D_r$	0.2

### 5.3 Training phase of the AHA–XDNN

At this phase, the three-layer DNN model was trained the hyper-parameter settings specified by AHA. With  $N_t = 100$ , the 3-layer DNN model employs the training set and validation set for training and assessment, respectively. To reduce the overfitting, the training procedure was pushed to end before  $N_t = K$  if no progress occurred after ten iterations. This dominance was carried out utilizing early stopping (Prechelt, 2012). Since the WQ dataset is a binary class classification problem, the three-layer DNN model was compiled using binary cross-entropy (Bosman et al., 2020). A step decay learning rate scheduler with initial value =  $1e - 3$  was applied to the Adam optimizer algorithm (Kingma & Ba, 2014; Senior et al., 2013).

### 5.4 Model evaluation and results explanation phases of AHA–XDNN

This section presents the AHA–XDNN approach’s outcomes. Accuracy, loss rate, precision, recall, and F1 score were utilized to gauge how well the proposed approach performed. The proposed AHA–XDNN achieved 91% accuracy on the test set. The average precision, recall, and F1-score for the proposed AHA\_XDNN on the test set are 91%, 91.5%, and 91%, respectively. As depicted in Table 5, the precision, recall, and F1-score had identical macro- and weighted average values of 91%.

The proposed AHA–XDNN approach’s ability to correctly classify test samples, as well as the number of examples it was unable to correctly classify, was both determined utilizing a confusion matrix, as depicted in Fig. 5. In the confusion matrix, samples that were successfully classified into each class are represented by the dark-colored shaded cells. Samples misclassified for each class are shown as light-colored shaded cells in the

**Table 5** Performance of the proposed approach AHA–XDNN on the test set

	Precision (%)	Recall (%)	F1-score (%)
Class 0	94	88	91
Class 1	88	95	91
Average	91	91.5	91
Macro-average	91	91	91
Weighted average	91	91	91

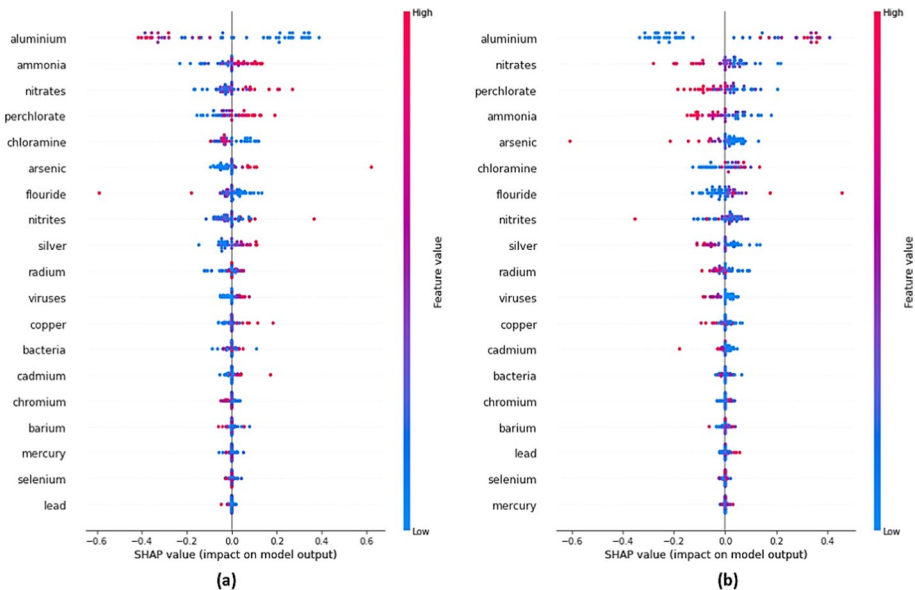
**Fig. 5** The confusion matrix generated by evaluating the AHA–XDNN approach on the test set

	Predicted Class_0	Predicted Class_1
True Class_0	T Positive= 270	F Negative= 38
True Class_1	F Positive= 16	T Negative= 276

confusion matrix. It should be noted that addressing the imbalance in the dataset helped a lot in preventing the proposed AHA–XDNN from biasing the once-dominant category. The proposed AHA–XDNN approach misclassified 54 samples from the test set, where it classified 36 samples from class 0 as being within class 1 and classified 16 samples from class 1 as being within class 0.

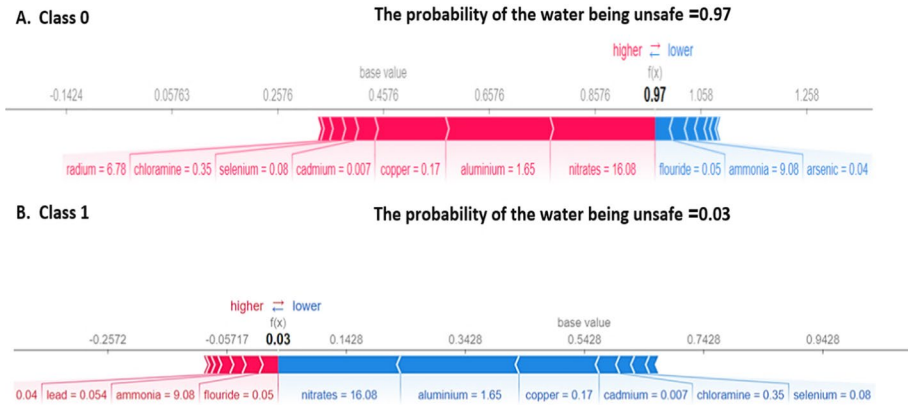
To interpret how the proposed AHA–XDNN determines to predict the test samples based on the input features' contribution to the model's performance, the SHAP analysis was used. As shown in Fig. 6, the SHAP summary plot demonstrates how each input feature impacts on the model performance. According to SHAP value, the input features in Fig. 6 are sorted so that a feature having the greatest influence on model's performance is displayed at a higher place. The colored points represent the SHAP value for each sample in the test set. The gradation of the colored points represents the data's actual value, ranging from low values (blue) to high values (red). In other words, positive SHAP values are expressed by the red colored points as can be seen in Fig. 6, while negative SHAP values are expressed by the blue colored points. The SHAP plot offered an explanation of how input features affect the predictions of AHA–XDNN approach. For instance, aluminum is at the top of the figure, indicating that aluminum had the most influence on the predictions. Therefore, from Fig. 6a, it can be explained that very low aluminum values tend to increase the prediction of the proposed approach for class 0. A higher value of ammonia was expressed by the red points on the right edge of Fig. 6a, indicating that the higher value of ammonia tends to increase the prediction of the proposed approach for class 0. On the contrary, a lower value of ammonia was represented by the blue points on the right edge of Fig. 6b, indicating that the lower value of ammonia tends to increase the prediction of the proposed approach for class 1.

SHAP analysis can also provide a detailed explanation of individual observations. The SHAP force plot can indicate exactly which features had the greatest influence on a model's



**Fig. 6** SHAP summary plot. **a** Represents SHAP summary for class 0 and **b** represents SHAP summary for class 1

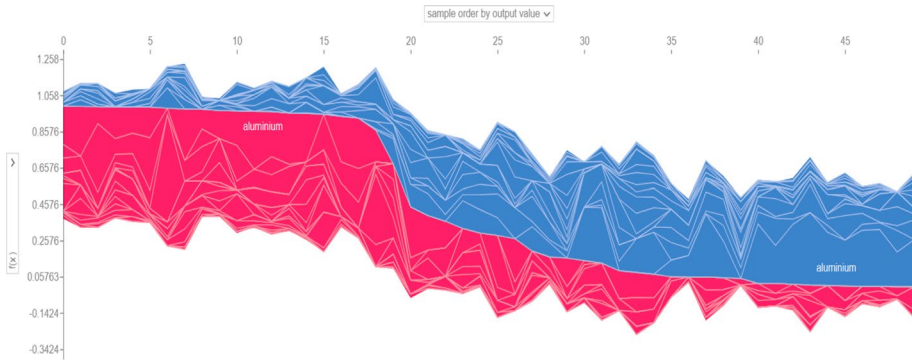




**Fig. 7** SHAP force plots. **A** SHAP force plot for a single observation of class 0 from the test set and **B** SHAP force plot for a single observation of class 1 from the test set

prediction for a single observation. This is ideal for being able to explain exactly how the model made a specific decision for a single observation. Figure 7 shows two SHAP force plots, one for each target class in the WQ test set. The binary target is class 0 which indicates that the water is not safe, and class 1 which indicates that the water is safe. Higher scores drive the model to predict class 0 and lower scores driving the model to predict 1. In Fig. 7, features that are crucial for making the prediction of the randomly selected observation are displayed in red and blue. Red denoting the features impulses the model’s prediction score higher. Blue denoting the features impulses the model’s prediction score lower. The features that had a greater effect on the score are situated near the red-blue splitter boundary, and the magnitude of this effect is represented by the bar’s size. Therefore, in Fig. 7a the proposed approach has been pushed to predict that the water is not safe (class 0) by the influence of the factors shown in red which are: nitrites, aluminum, copper, cadmium, selenium, chloramine, and radium, respectively. On the other side, in Fig. 7b the proposed approach has been pushed to predict that the water is safe (class 1) by the influence of the factors shown in blue which are: nitrites, aluminum, copper, cadmium, chloramine, and selenium, respectively. If the force plots of all the observations in the test set are combined, rotated 90°, and stacked horizontally, the combined force plot as shown in Fig. 8 will be obtained. As shown in Fig. 8, aluminum is the most influential feature in most but not all predictions.

The AHA–XDNN was contrasted with other published models that were introduced for the same aim of predicting WQ using the same dataset as utilized in this paper to ensure its performance. After examining all the literature reviewed in Sects. 1 and 2, it was found that there was only one model for predicting WQ using the same dataset as utilized in this paper (Rustam et al., 2022). In (Rustam et al., 2022), accuracy, precision, recall, and F1-score were the four metrics utilized to assess the proposed ANN model. As depicted in Table 6, the ANN model achieves higher accuracy than the proposed approach. But this is because the ANN model has been trained on the WQ dataset without being treated from the imbalance issue. Since accuracy is not a good measure if the dataset is unbalanced, it is better to compare the AHA–XDNN’s performance and ANN model in terms of F1-score, precision, and recall measurements. These measurements should be high to indicate perfect classifier’s performance. ANN model and the proposed approach AHA–XDNN achieved an equal



**Fig. 8** SHAP force plot across all observations in the test set

**Table 6** Performance assessment of the AHA–XDNN approach in comparison with existing models

Performance measurement	Proposed AHA–XDNN (%)	ANN model (Rustam et al., 2022) (%)
Accuracy	91	96
Average precision	91	91
Average recall	91.5	87
Average F1-score	91	89

average precision rate of 91%. ANN model achieved a low recall rate of 87% which means it has a higher number of false-negative predictions (water is safe but was incorrectly predicted as not safe). The proposed approach outperformed the ANN model, achieving an average recall rate of 91.5% and an average F1-score rate of 91%.

## 6 Conclusion and future work

Water is pivotal to sustainable development since it is required for social and economic growth, healthy ecosystems, and human life. WQ is vital to society and ecology, making it a significant aspect in reaching the SDGs. This paper presents an XAI approach called AHA–XDNN for predicting WQ. The proposed approach is split into five phases. The first is data preparation, which addresses issues in the used dataset such as undesired noise and imbalance. The second phase is the optimization phase, in which AHA is implemented to select the ideal values for the DNN model’s hyper-parameters, which have a significant influence on its performance. The DNN model, optimized using AHA, is trained on the dataset in the third phase, which is the training phase. In the fourth phase, four measurements are utilized to measure how well the optimized DNN model performs. These measurements are accuracy, recall, precision, and F1-score. On the test set, the proposed AHA–XDNN accomplished a competitive accuracy level of 91%. The results of the optimized DNN model are explained using the most common XAI technique “SHAP” in phase fifth. SHAP measures the interplay between each parameter and the contribution to the final result, convincing the fundamental nature of the ML models. Furthermore, SHAP

offers the causality and an interpretation of the inner workings of proposed approach to increase end-user trust in AHA–XDNN’s decisions. These recent breakthroughs in interpretable ML allow us to see inside the black box and explain how each observation’s prediction works. Despite the promising results of DL-based methods in predicting WQ, the large number of features can negatively affect the performance of classifiers and make them more complex. As a future direction, the SHAP might be employed as a feature reduction tool, which would increase the accuracy of the proposed approach while reducing computational costs.

**Funding** Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB).

**Data availability** Data that support the findings of this study have been deposited in [Kaggle: <https://www.kaggle.com/datasets/mssmartypants/water-quality>].

## Declarations

**Conflict of interest** There is no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abdulla, A.F. (2021). Bacterial diseases caused by contaminated drinking water
- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, *6*, 52138–52160.
- Awad, M., & Khanna, R. (2015). Deep neural networks. In M. Awad & R. Khanna (Eds.), *Efficient learning machines*. Apress. [https://doi.org/10.1007/978-1-4302-5990-9\\_7](https://doi.org/10.1007/978-1-4302-5990-9_7)
- Bi, J., Lin, Y., Dong, Q., Yuan, H., & Zhou, M. (2021). Large-scale water quality prediction with integrated deep neural network. *Information Sciences*, *571*, 191–205.
- Bosman, A. S., Engelbrecht, A., & Helbig, M. (2020). Visualising basins of attraction for the cross-entropy and the squared error neural network loss functions. *Neurocomputing*, *400*, 113–136.
- Carneiro, T., Medeiros Da Nóbrega, R. V., Nepomuceno, T., Bian, G., De Albuquerque, V. H. C., & Filho, P. P. R. (2018). Performance analysis of google colabatory as a tool for accelerating deep learning applications. *IEEE Access*, *6*, 61677–61685.
- Charles, J., Vinodhini, G., & Nagarajan, R. (2021). An efficient feature selection with weighted extreme learning machine for water quality prediction and classification model. *Annals of the Romanian Society for Cell Biology*, *25*, 1969–1994.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. <https://doi.org/10.1613/jair.953>
- Chollet, F. (2015). Keras: Deep learning for humans [Online]. <https://github.com/fchollet/keras>.
- Darwish, A., Ezzat, D., & Hassanien, A. E. (2020). An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. *Swarm and Evolutionary Computation*, *1*(52), 100616.
- Delanka-Pedige, H. M., Munasinghe-Arachchige, S. P., Abey Siriwardana-Arachchige, I. S., & Nirmalakhandan, N. (2021). Wastewater infrastructure for sustainable cities: Assessment based on UN sustainable development goals (SDGs). *International Journal of Sustainable Development and World Ecology*, *28*(3), 203.

- Farhi, N., Kohen, E., Mamane, H., & Shavitt, Y. (2021). Prediction of wastewater treatment quality using LSTM neural network. *Environmental Technology and Innovation*, 23, 101632.
- Garcia, J. C., Robb, D. A., Liu, X., Laskov, A., Patron, P., & Hastie, H. (2018). Explain yourself: A natural language interface for scrutable autonomous robots. In *Proceedings of explainable robotic system workshop HRI*.
- Gohel, P., Singh, P., & Mohanty, M. (2021). Explainable AI: current status and future directions.
- Goutte, C., & Gaussier, E. (2005) A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *Proceedings of European conference on information retrieval* (pp. 345–359). Springer.
- Hassan, M. M., Hassan, M. M., Akter, L., et al. (2021). Efficient prediction of water quality index (WQI) using machine learning algorithms. *Human -Centric Intelligent Systems*, 1(3–4), 86–97.
- Ioffe, S., & Szegedy C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. <https://doi.org/10.48550/ARXIV.1502.03167>
- Jahangard, H., Hosseini, S. M., Ebadi, T., & Akbarimehr, D. (2022). Effect of gas oil contamination on dynamic geotechnical properties of clay-sand mixtures. *Arabian Journal of Geosciences*, 15(3), 1–16.
- Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *Journal of Big Data*, 6, 27. <https://doi.org/10.1186/s40537-019-0192-5>
- Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A., Mohamed, A., & Ashraf, I. (2022). Water quality prediction using KNN imputer and multilayer perceptron. *Water*, 14, 2592. <https://doi.org/10.3390/w14172592>
- Khan, Y., & See, C. S. (2016). Predicting and analyzing water quality using machine learning: A comprehensive model. In *Proceedings of the 2016 IEEE long Island systems, applications and technology conference (LISAT)*, April 2016.
- Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T., & Thuy, N. T. D. (2022). Using machine learning models for predicting the water quality index in the La buong river, Vietnam. *Water*, 14(10), 1552.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization [Online]. [arXiv:1412.6980](https://arxiv.org/abs/1412.6980)
- Liang, Y., Ke, S., Zhang, J., Yi, X., & Zheng, Y. (2018). GeoMAN: Multi-level attention networks for geo-sensory time series prediction. In *Proceedings of 27th international joint conference on artificial intelligence* (pp. 3428–3434).
- Lundberg, S. & Lee, S.-I. (2017). A unified approach to interpreting model predictions. <https://doi.org/10.48550/ARXIV.1705.07874>
- Madni, H. A., Umer, M., Isha, A., Abuzinadah, N., Saidani, O., Alsubai, S., Hamdi, M., & Ashraf, I. (2023). Water-quality prediction based on H<sub>2</sub>O AutoML and explainable AI techniques. *Water*, 15, 475. <https://doi.org/10.3390/w15030475>
- Nair, J. P., & Vijaya, M. S. (2021). Predictive models for river water quality using machine learning and big data techniques: A Survey. In *Proceedings of the 2021 international conference on artificial intelligence and smart systems (ICAIS)*, IEEE, Coimbatore, India.
- Neerinx, A., van der Waa, J., Kaptein, F., & van Diggelen, J. (2018). Using perceptual and cognitive explanations for enhanced human-agent team performance. In *Proceedings of international conference engineering psychology and cognitive ergonomics (EPCE)* (pp. 204–214).
- Nhamo, G., Nhemachena, C., & Nhamo, S. (2019). Is 2030 too soon for Africa to achieve the water and sanitation sustainable development goal? *Science of the Total Environment*, 15(669), 129–139.
- Panigrahi, N., Patro, S. G. K., Kumar, R., et al. (2023). Groundwater quality analysis and drinkability prediction using artificial intelligence. *Earth Science Informatics*, 16, 1701–1725. <https://doi.org/10.1007/s12145-023-00977-x>
- Park, J., Ahn, J., Kim, J., Yoon, Y., & Park, J. (2022a). Prediction and interpretation of water quality recovery after a disturbance in a water treatment system using artificial intelligence. *Water*, 14, 2423. <https://doi.org/10.3390/w14152423>
- Park, J., Lee, W. H., Kim, K. T., Park, C. Y., Lee, S., & Heo, T.-Y. (2022b). Interpretation of ensemble learning to predict water quality using explainable artificial intelligence. *Science of the Total Environment*, 832, 155070. <https://doi.org/10.1016/j.scitotenv.2022.155070>
- Patel, J., Amipara, C., Ahanger, T., Ladhva, K., Gupta, R., Alsaab, H., Althobaiti, Y., & Ratna, R. (2022b). A machine learning-based water potability prediction model by using synthetic minority oversampling technique and explainable AI. *Computational Intelligence and Neuroscience*, 2022, 1–15. <https://doi.org/10.1155/2022/9283293>
- Prechelt, L. (2012). Early stopping—But when? In G. Montavon, G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (2nd ed., pp. 53–67). Springer.
- Qin, Y., Song, D., Chen, H., Cheng, W., Jiang, G., & Cottrell, G. (2017). A dual-stage attention-based recurrent neural network for time series prediction. [arXiv:1704.02971](https://arxiv.org/abs/1704.02971)

- Rasheed Abdul Haq, K. P., & Harigovindan, V. P. (2022). Water quality prediction for smart aquaculture using hybrid deep learning models. *IEEE Access*, *10*, 60078–60098. <https://doi.org/10.1109/ACCESS.2022.3180482>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?”: Explaining the predictions of any classifier. <https://doi.org/10.48550/ARXIV.1602.04938>
- Rustam, F., Ishaq, A., Kokab, S. T., de la Torre, D. I., Mazón, J. L., Rodríguez, C. L., & Ashraf, I. (2022). An artificial neural network model for water quality and water consumption prediction. *Water*, *14*(21), 3359.
- Senior, A., Heigold, G., Ranzato, M., & Yang, K. (2013). An empirical study of learning rates in deep neural networks for speech recognition. In *Proceedings of IEEE international conference on acoustics, speech and signal processing* (pp. 6724–6728).
- Shapley, L. S. (1953). A value for n-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games 2.28* (pp. 307–317). Princeton University Press.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(56), 1929–1958.
- Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*, *17*, 168.
- Ting, K. M. (2011). Confusion matrix. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of machine learning* (p. 209). Springer.
- VenkataVara Prasad, D., Venkataramana, L. Y., Senthil Kumar, P., Prasannamedha, G., Harshana, S., JahnaviSrividya, S., Harrine, K., & Indraganti, S. (2022). Analysis and prediction of water quality using deep learning and auto deep learning techniques. *Science of the Total Environment*, *821*, 153311. <https://doi.org/10.1016/j.scitotenv.2022.153311>
- Wang, Y., Zhou, J., Chen, K., Wang, Y., Liu, L. (2017). Water quality prediction method based on LSTM neural network. In *November 2017 12th international conference on intelligent systems and knowledge engineering (ISKE)* (pp. 1–5). IEEE.
- Wang, Z., Wang, Q., & Wu, T. (2023). A novel hybrid model for water quality prediction based on VMD and IGOA optimized for LSTM. *Frontiers of Environmental Science and Engineering*, *17*, 88. <https://doi.org/10.1007/s11783-023-1688-y>
- Yahya, A., Najah, A.-M., Othman, F., Ibrahim, R., Afan, H., El-Shafie, A., Chow, M. F., Hossain, Md., Ehteram, M., & Elshafie, A. (2019). Water quality prediction model based support vector machine model for ungauged river catchment under dual scenarios. *Water*, *11*, 1231. <https://doi.org/10.3390/w11061231>
- Zhao, J., Huang, F., Lv, J., Duan, Y., Qin, Z., Li, G., & Tian, G. (2020). Do RNN and LSTM have long memory. In H. D. III, & A. Singh (Eds) *Proceedings of 37th international conference on machine learning* (vol. 119, pp. 11365–11375).
- Zhao, W., Wang, L., & Mirjalili, S. (2022). Artificial hummingbird algorithm: A new bio-inspired optimizer with its engineering applications. *Computer Methods in Applied Mechanics and Engineering*, *388*, 114194. <https://doi.org/10.1016/j.cma.2021.114194>
- Zheng, W., Zhao, P., Huang, K., & Chen, G. (2021) Understanding the property of long term memory for the LSTM with attention mechanism. In *Proceedings of the 30th ACM international conference on information and knowledge management* (pp. 2708–2717).
- Zhou, M., Huang, M., & Zhu, X. (2018). An interpretable reasoning network for multi-relation question answering. In *Proceedings of the 27th international conference on computational linguistics*.
- Zhou, J., Wang, Y., Xiao, F., Wang, Y., & Sun, L. (2018). Water quality prediction method based on IGRA and LSTM. *Water*, *10*, 1148.