IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 5, 2021, accepted April 18, 2021, date of publication April 26, 2021, date of current version May 6, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3075849

Hybrid Predictive Model for Water Quality Monitoring Based on Sentinel-2A L1C Data

GEHAD HASSAN[1], MOHAMED E. GOHER[2], MASOUD E. SHAHEEN[1], AND SHEREEN A. TAIE[1]
[1]Computer Science Department, Faculty of Computers and Information, Fayoum University, Fayoum 63514, Egypt
[2]National Institute of Oceanography and Fisheries (NIOF), Cairo 11694, Egypt

Corresponding author: Gehad Hassan (gha11@fayoum.edu.eg)

ABSTRACT Monitoring water quality is an important challenge in both developed and developing countries. Remote sensing data can form a highly frequent dataset with acceptable spatial coverage that can be used to remotely monitor water quality. This paper presents a novel automated model for remotely monitoring water quality to address the problem of insufficient samples and save the time and cost of sample collection. The proposed model estimates both optical and non-optical water quality parameters via Sentinel-2A data. A bio-inspired hybrid model of a Binary Whale Optimization Algorithm (BWOA) and Artificial Neural Network (ANN) (BWOA-ANN) is applied to determine the relationship between extracted reflectance values from Sentinel-2A images and analyzed samples. The novelty of this model is to solve two main problems of remote water quality monitoring: poor applicability and low non-optical parameter estimation accuracy. For the first problem, a proposed fully automated model with band selection using the BWOA to automatically select the optimal features (Sentinel-2A bands) that are suitable for each water quality parameter. The second problem is addressed by automatically detecting the relationship between non-optical parameters, such as the total phosphorus, and optical parameters, such as chlorophyll-a. Three datasets with different locations, seasons, and parameters were selected to test the proposed BWOA-ANN. The experimental results demonstrated good regression with a mean R^2 value of 0.916 for optical parameters and 0.890 for non-optical parameters. The proposed model was found to outperform the ANN with an R^2 value higher by 40% and 52% for the optical and non-optical parameters, respectively.

INDEX TERMS Artificial Neural Network (ANN), feature selection, Sentinel-2, Whale Optimization Algorithm (WOA), water quality monitoring.

I. INTRODUCTION

Water is one of the most vital elements for the survival of humans, plants, and animals. Water quality has the same importance as water itself because of its key role in sustainable development [1], human health, fishing habits, and human food [2]–[5]. Contaminated water consumption is the cause of approximately 502,000 global deaths each year, mostly in economically challenged countries [6]. Many factors affect the water quality, such as that caused by agricultural and industrial residues, and anthropogenic activities [5]. Therefore, at the beginning of the 20th century, water quality started being taken more seriously [2]. Many studies have been tried to investigate mechanisms to improve water quality [7]–[10]. This, in turn, requires continuous monitoring of water quality to assess these methods and test their effectiveness, demonstrating the need for a low-cost and easy-to-use yet robust and scientifically proven method for determining water quality.

Taking samples from different locations along the water surface and analyzing them in a laboratory is the common way to assess water quality parameters [4], [11]. Although in situ measurements provide accurate results, they require a large amount of time and cost [5], are labor-intensive, and do not provide satisfactory results on temporal and spatial ranges [12]. Furthermore, water quality is not constant along the water surface [13] or in different seasons [4], [13]. These limitations are thus significant obstacles for water quality assessment using in situ measurement methods [14].

Remotely sensed data has the capability of covering large spatial scale with frequently revisit times [2], [15]. Consequently, this advantages make remote sensing data considered

The associate editor coordinating the review of this manuscript and approving it for publication was Zijian Zhang.

as a good repository for environmental condition monitoring, such as soil moisture estimation [16], precipitation estimations [15], water level monitoring [17], water quality monitoring [2], and flood mapping [18]. For water quality monitoring control, satellite images can cover large spatial and temporal scales, which cannot be accomplished by in situ sampling.

According to remote sensing, water quality parameters can be classified as optical parameters and non-optical parameters. Optically active parameters include Chlorophyll-a (Chl-a), Total Dissolved Solids (TDS), Temperature (Temp), Transparency (Trans), Total Suspended Solids (TSS), and Turbidity [2], [19], [20]. The non-optical water quality parameters include Dissolved Oxygen (DO), Chemical Oxygen Demand (COD), Nitrate (NO_3), Phosphate (PO_4), Total Phosphorus (TP), and Total Nitrogen (TN) [2], [19], [21], [22]. It should be noted that most of the studies focused on optically active parameters and detected non-optical parameters with low accuracy [23]–[26]. However, a number of vital water quality parameters, such as TN, TP, PO_4, NO_3, and ammonia nitrogen ($NH_3 - N$), have not been thoroughly examined due to their poor optical features, which is a key challenge in remote water quality monitoring [25], [27].

Many researchers widely use classical regression analysis to estimate water quality parameters. The performance of these methods is strongly affected by the variation of season, case study, and investigated parameters [2] as the optical properties of local water sources are not stable at all times [2]. Usually, these studies manually define the best bands for a specific water source for a certain season, so it is not applicable for any other water quality parameter or another water source. Accordingly, to solve this problem and improve the performance of water quality parameter estimation, there is a need to build a generalized model that can be automatically updated according to the case study and the investigated parameters.

Feature selection is an important step in the preprocessing phase. It aims to select the best features that positively affect the model performance [28]. The brute-force method is not a good solution for feature selection tasks as it leads to high computational time. Also, manually selecting these optimal features is not a practical solution. Recently, many meta-heuristic algorithms such as Genetic Algorithm (GA), Grey Wolf Optimization (GWO), Cuckoo-Search Algorithm (CSA), Whale Optimization Algorithm (WOA), Bat Algorithm (BA), etc. are efficiently used as solutions for feature selection [28], [29]. Accordingly, in this paper, to improve the estimation of optical and non-optical water quality parameters, we designed a band selection process, based on Binary Whale Optimization Algorithm (BWOA).

The aim of the current research is to develop a novel predictive bio-inspired hybrid model to estimate both optical and non-optical water quality parameters with accurate results using Sentinel-2A data. An automatic band selection process based on Binary Whale Optimization Algorithm (BWOA) and Artificial Neural Network (ANN) (BWOA-ANN) is designed to improve regression accuracy by automatically selecting the appropriate Sentinel-2A bands for each water quality parameter. This process allows the model to be applied to different case studies, and seasons. Poor estimation accuracy is avoided for non-optical parameters by automatically detecting the relationship between non-optical and optical parameters. The applicability of the proposed model is tested and validated on different real-world datasets (obtained by in situ sampling) for different seasons and locations.

The remainder of this paper is organized as follows. Section (II) presents state-of-the-art studies related to water quality parameters monitoring using remote sensing data. Section (III) describes the selected locations for this study, the in situ data, and the selected sensor. Section (IV) explains an overview about the methods used in the proposed model. Then, Section (V) describes the different phases of the proposed model. Section (VI) discusses the experimental results, and Section (VII) presents the conclusions.

II. RELATED LITERATURE

Previously, many studies used Landsat TM [2] to retrieve water quality parameters. The long repeated temporal range and reduced radiometric resolution of Landsat TM restrict it from being used for water quality monitoring, however [2], [30]. Recently, studies have shown the utility of Landsat-8 [19], [21], [22], [31]–[33], Sentinel-2 [19], [22], [24], [32]–[35], and Sentinel-3 [25], [32], [36]. Sentinel-2 and Sentinel-3 present an excellent opportunity to monitor optical water quality parameters over the water source owing to the fine spatial resolution for the first sensor and the fine spectral resolution for the second sensor, high data quality, and availability [24], [32], [34]. Additionally, in [37], Buma et. al differentiated between Sentinel-2 and Landsat-8 capabilities and confirmed that Sentinel-2 overcomes Landsat-8 in some advantages. Sentinel-2 has more fine temporal resolution and spectral resolution than Landsat-8. Also, the authors illustrated that Sentinel-2 has a lower Signal to Noise Ratio (SNR) compared to Landsat-8 [37].

According to the literature, many studies have also applied band ratio algorithms using Sentinel-2 to analyze and define the appropriate bands for some specific water quality parameter [38], [39]. Brezonik, *et al.* [32] used Sentinel-2, Sentinel-3, and Landsat-8 data to study factors that have a high impact on Colored Dissolved Organic Matter (CDOM) estimation including spectral resolution and season variation. The main advantage of this study is to deeply analyze various long term datasets. The authors noted that CDOM concentration visually varies in (400–700 nm) among surface waters and by the time in a given water body. This study, unfortunately, aimed to define which specific bands are suitable for estimating CDOM, so it is not applicable for any other water quality parameter or another case study.

Additionally, Yuchao *et al.* [40] tried to work on the challenge of large turbidity ratio and big biomass in shallow lakes in China. The authors performed Baseline Normalized Difference Bloom Index (BNDBI) using Band4 and Band1 of Moderate Resolution Imaging Spectroradiometer (MODIS)

data. BNDBI has an advantage that the absorption of Chl-a near to 572 nm wavelength is minimum and near to 667 nm wavelength is maximum. BNDBI is characterized as a vital approach to monitor algal progress but within the optical scope (aquatic and atmospheric). The authors continue their work in [41] using a different algorithm that uses the correlation between surface Chl-a and algal biomass of water at different depths to calculate the total algal biomass under the non-algae bloom conditions. Validation sites proved that the proposed approach is consistent with in-situ samples and stable. It was observed that the performance of this algorithm reduced the Root Mean Square Error (RMSE) from 77.9% (in case of using BNDBI) to 32.5%. Also, Ha *et al.* [34] are ones of those who investigated specific bands for Chl-a estimation using Sentinel-2A data and in-situ measurements. The authors achieved R^2 values of 0.68 using 560 nm and 665 nm wavelengths and R^2 values of 0.29 using 705 nm and 665 nm wavelengths for Vietnams Lake Ba Be. Moreover, Masocha *et al.* in [21], investigated Landsat-8 data capability against estimating some of the water quality parameters for tow lakes in Zimbabwe. The authors concluded that the Near-infrared/Red band was suitable for Chl-a estimation with the less polluted lakes, and the red band was suitable for the more polluted one.

In [24], Karaoui *et al.* built a relationship between in-situ measurements and Sentinel-2 data reflectance values. The authors aimed to prove that Sentinel-2 could be used as a good tool for water quality monitoring. Stepwise regression achieved R^2 values greater than 0.52 for all of Chl-a, DO, NO_3, PO_4, and TP. It is worth noting that non-optical parameters had lower values of R^2.

Most of the previously mentioned studies make a classical regression analysis that does not achieve satisfactory accuracy because of the complexity of the nonlinear relations of the problem. However, intelligent models -such as Support Vector Machine SVM, ANN, and Fuzzy Logic- offer a good solution to deal with these complex systems. Regarding this, Batur and Maktav [19] applied Principal Component Analysis (PCA) data fusion using Landsat-8 Operational Land Imager (OLI), Sentinel-2A, and Gokturk-2 data to estimate Chl-a, DO, TSS, SDD, TDS, and pH values in Lake Gala, Turkey. Moreover, the authors compared multiple linear regression, ANN, and SVM results which were applied to the three sensors data. It was demonstrated that PCA results were the most correlated with in-situ measurements achieving R^2 about 0.92, 0.89, 0.89, 0.88, 0.73, and 0.88 for Chl-a, TSS, DO, SDD, TDS, and pH, respectively.

In [36], Blix *et al.* tested the capability of Sentinel-3 Ocean and Land Color Instrument (OLCI) to monitor optically active water quality parameters (Chl-a, CDOM, and Total Suspended Matter (TSM)). The authors used the Automatic Model Selection Algorithm (AMSA) [42] to rank the spectral bands and combine the most related bands to Chl-a retrieval. Gaussian Process Regression (GPR) was applied for each combination of features. The previous step was applied for each station in the observed lake to guarantee that GPR would not be spectrally affected by water conditions changes. The model selected Band3 and Band5 as the appropriate bands to estimate Chl-a concentration with an R^2 value of 0.82. It was observed that the proposed AMSA GPR model performance was decreased when all Sentinel-3 OLCI remote sensing bands were added, achieving an R^2 value of 0.7909.

The neural network has great potential to be used for water quality parameter estimation via remote sensing. ANN was indirectly used in [43], [44] to improve Chl-a retrieval for both open and complex water sources via atmospheric correction. Also, Govedarica and Jakovljevic [33] used ANN and SVM to analyze the relationship between in-situ measurements and the reflectance values of images. The authors used Sentinel-2 and compared it with Landsat-8 results. It was illustrated that Landsat-8 needed atmospheric correction before applying the model that pushed the authors to give the advantage to Sentinel-2 to estimate TP and TN. Du *et al.* in [25] improved the accuracy of TP concentration estimation-as non-optical parameter-by applying a new optical classification method that classified water into two different types. The authors applied Data Regression Analysis and Fitting (DRF), Backpropagation Neural Network (BPNN), and Random Forest (RF) on Sentinel-3 data. The authors indicated that DRF performed best results with an R^2 value of 0.80 and an RMSE value of 0.034. In a similar vein, Sharaf El-Din *et al.* in [45], proposed a framework based on the BPNN, but the authors in this work used Landsat-8 data. The BPNN results were compared with SVM, BPNN outperformed SVM with R^2 values of 0.991, 0.933, 0.937, 0.930, and 0.934 for turbidity, TSS, COD, BOD, and DO, respectively.

Other significant literature tried to apply Case 2 Regional CoastColour (C2RCC) to measure water quality parameters via remote sensing [46]. C2RCC uses a neural network and has been trained on extreme ranges of scattering and absorption properties. C2RCC is applicable for almost of ocean color sensors as well as Sentinel-2 and Sentinel-3 [46], [47]. The C2RCC processor contains two main parts: the atmospheric correction part, and the in-water part. In [47], the authors tested C2RCC with Sentinel-3A OLCI data to retrieve Inherent Optical Properties (IOPs), and some of the optical water quality parameters from the Baltic sea. It was concluded that C2RCC performed reasonable results in case of atmospheric correction, but it did not perform well in retrieving water quality parameters. The authors demonstrated that the neural network in C2RCC needs more training with the actual data to output accurate results. Also, Ansper *et al.* [48] tested the standard C2RCC in Chl-a estimation but with Sentinel-2 data. The authors also ensured that the C2RCC processor needs more validation data to provide reasonable results and investigated that C2RCC results are not accurate with small and narrow lakes. Sidrah *et al.* [49] compared C2RCC with machine Learning algorithms for the retrieval of optical water quality parameters. The applied ANN in [49] outperformed the C2RCC results in terms of RMSE. Consequently, many studies confirmed that C2RCC

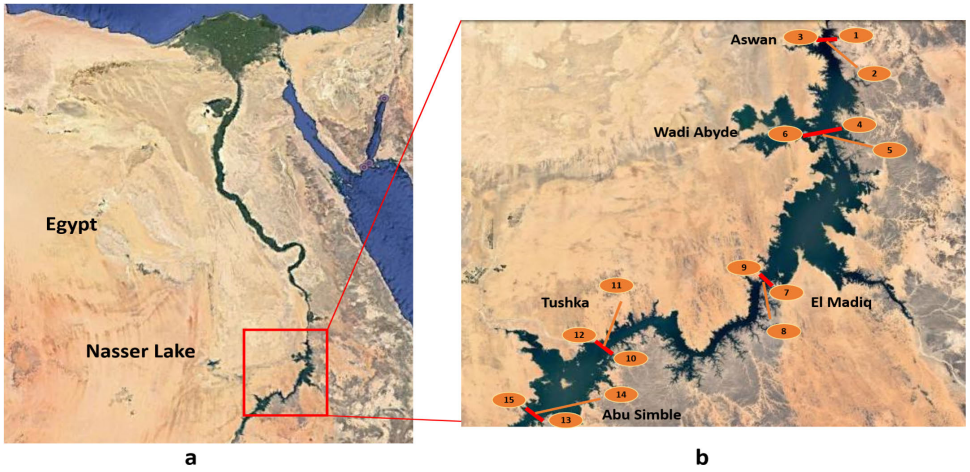

FIGURE 1. Study area 1: Nasser Lake and locations of in situ measurements taken on August 2016.

needs more training [46], [47] and does not perform well in most cases of water quality parameters retrieval [49], [50].

Deep learning well performs with the most remote sensing problems [51], but there are rarely studies that used it for water quality monitoring. A novel model that used deep learning is proposed in [22]. The authors developed progressively decreasing deep neural networks and implement it on both Landsat-8 and Sentinel-2 data. The proposed model in [22] outperforms Multiple Linear Regression (MLR), Support Vector Machine Regression (SVR), and Extreme Learning machine Regression (ELR) with the most popular remotely investigated water quality parameters. The resulting R^2 were 0.839, 0.926, 0.894, and 0.883 for turbidity, Fluorescent Dissolved Organic Matter (fDOM), DO, and Chl, respectively. It was observed that the performance of the proposed water quality estimation model is not stable against different optical and non-optical parameters. However, deep learning is more powerful with the various applications of remote sensing, there are still some limitations that lead to extra challenges of using deep learning with remote sensing [27] because of the lack of used in-situ and remote sensing data.

III. MATERIALS

This section presents an overview of the study areas, in situ data, and Sentinel-2 data.

A. STUDY AREAS AND IN SITU DATA

The proposed model is applied based on real datasets, namely, two datasets for different seasons for Nasser Lake, Egypt, and a dataset for the Bin El Ouidane Reservoir, Morocco [24]. Two different water sources were selected with diverse environmental conditions to prove that the model well performed against various water sources.

1) STUDY AREA 1 (NASSER LAKE, AUGUST 2016)

Nasser Lake is considered the main strategic fresh water storage in Egypt with over 95% of the total Egyptian freshwater proceeds [52], [53], and is deemed the second largest lake in the world [52]. This lake extends from the southern part of Egypt to the northern part of Sudan [53]. It has a length of approximately 500 km, of which 350 km are in Egypt. The total area of Nasser Lake is approximately 6,276 km^2, of which 5,237 km^2 are in Egypt at 180 m over the mean sea level [53]. Chl-a is the dominating of the watercolor for the lake. TSM is the more effective during the flood period (late July-October) in the southern sector. Chl-a mainly absorbs the energy from wavelengths of red and blue lights and reflects the green light, causing Chl-a to be appeared as green [2], [30]. The lake's bottom can not be seen. Water depth of all surface samples are illustrated in Table 2. Moreover, Aerosol concentrations increase during the summer and autumn due to the northward extension of the Sudan monsoon trough. Study area 1 of the proposed model was defined by a bounding box with coordinates between $22°00' - `23°58'N$ and $31°19' - 33°15'E$ of Nasser Lake, as illustrated in Figure 1(a). Water samples were collected in August 2016 from five sectors along the main channel, including the Aswan, Wadi Abyad, El Madiq, Tushka, and Abu-Simbel sectors. Three sites were selected in each sector at the eastern and western banks in addition to the main channel for a total of 15 sites (see Fig. 1b). Subsurface water samples were collected using a 2L Ruttner water sampler bottle that was kept in well-cleaned plastic bottles in an icebox. Water temperature (0 C), pH, and conductivity (EC, μs/cm) were measured in situ using the Hydrolab model (Multi Set 430i WTW) after applying a calibration procedure. Transparency was measured using a Secchi disk with a diameter of 30 cm. The chemical parameters of the water samples were determined using the standard methods of American Public Health Association (APHA) [54].

2) STUDY AREA 2 (NASSER LAKE, APRIL 2016)

The second study area was defined by a bounding box with coordinates between $21°54' - `23°58'N$ and $30°00' - 33°15'E$

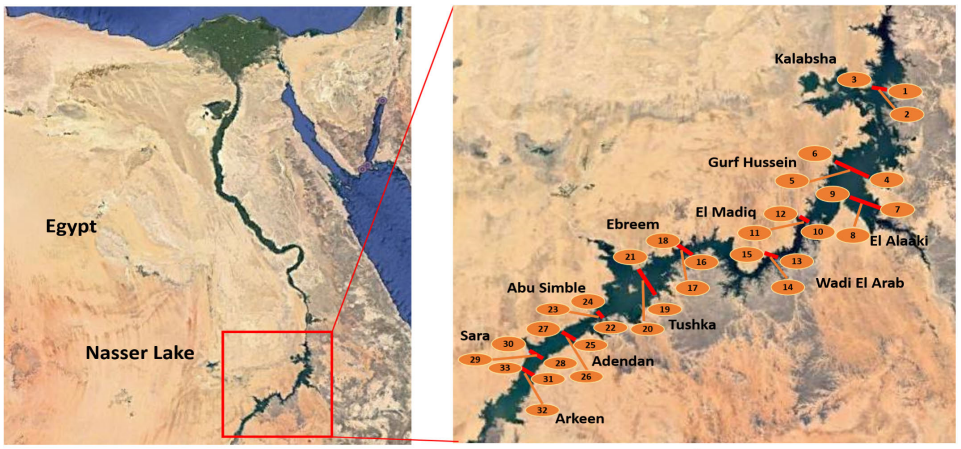

FIGURE 2. Study area 2: Nasser Lake and locations of in situ measurements taken on April 2016.

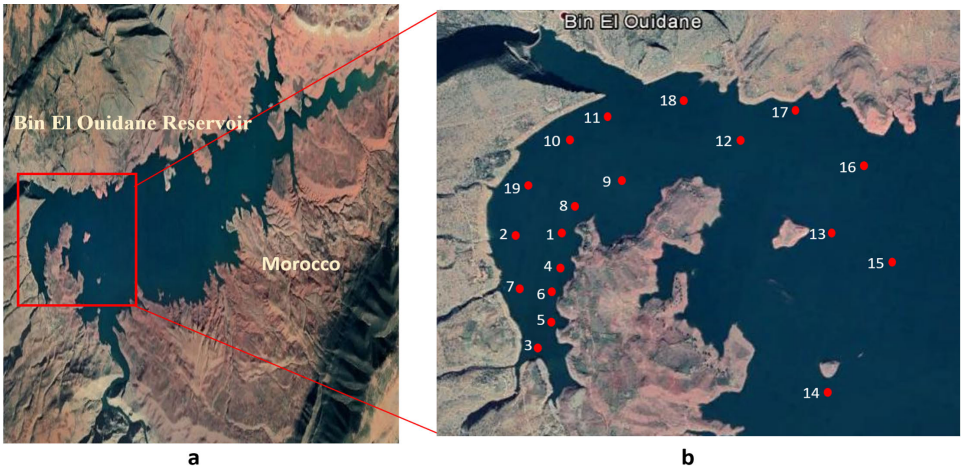

FIGURE 3. Study area 3: Bin El Ouidane Reservoir and locations of in situ measurements taken on May 2017.

of Nasser Lake, as illustrated in Figure 2(a). Water samples were collected in April 2016 from 11 sectors along the main channel, including the Kalabsha, Gurf Hussein, El Alaaki, El Madiq, Wadi El Arab, Ebreem, Tushka, Abu-Simbel, Adendan, Sara, and Arkeen sectors. Three sites were selected in each sector at the eastern and western banks in addition to the main channel for a total of 33 sites (see Fig. 2(b)). Water samples of study area 2 were collected with the same method as used for study area 1.

3) STUDY AREA 3 (BIN EL OUIDANE RESERVOIR, MAY 2017)

The Bin El Ouidane Reservoir is located in the center of the northern portion of Morocco in the Azilal area (Tadla Azilal region). This reservoir is the third largest reservoir in Morocco in terms of stored water [24]. The total area of the Bin El Ouidane Reservoir is approximately 36 km^2, and the reservoir contains 1,384 million m^3 of water [24]. Study area 3 was defined by a bounding box with coordinates between $32°88' - '32°92'N$ and $6°43' - 6°45'E$ of the Bin El Ouidane Reservoir, as illustrated in Figure 3(a).

Additional information on this study area is available in [24]. A total of 19 samples were collected downstream of the reservoir in several hours owing to climate fluctuations (see Fig. 3(b)).

B. SENTINEL-2 DATA

Sentinel-2A Level-1C (L1C) data were obtained from the Sentinels Scientific Data Hub (https://scihub.copernicus.eu/). Sentinel-2A L1C images consisted of 100 km^2, each with a volume of approximately 500 MB. These images underwent geometric and radiometric correction including orthorectification. The Sentinel-2A L1C images were resampled with a constant ground sampling distance of 10, 20, and 60 m depending on the resolution of the various spectral bands (see Table 1) [55]. The Sentinel Application Platform (SNAP) version 6.0.0 with the Sentinel-2 Toolbox (S2TBX) version 6.0.4 Windows 10 (64-bit) was used to process the Sentinel-2 images.

IV. PRELIMINARIES

A. ARTIFICIAL NEURAL NETWORKS

ANNs can learn, recognize, and handle various complex tasks in engineering and science by considering examples without

TABLE 1. Spectral bands, central wavelengths (nm), bandwidths (nm), and corresponding spatial resolutions (m) of Sentinel-2A L1C sensor.

Band Number	Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution (m)
1	443	20	60
2	490	65	10
3	560	35	10
4	665	30	10
5	705	15	20
6	740	15	20
7	783	20	20
8a	842	115	10
8b	865	20	20
9	945	20	60
10	1375	30	60
11	1610	90	20
12	2190	180	20

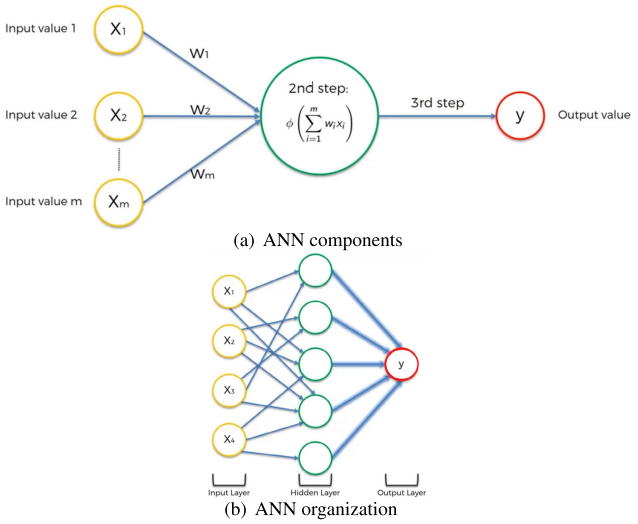

FIGURE 4. Artificial Neural Network (ANN).

needing to be programmed using task-specific rules [56]. ANNs are effective learning models that can produce satisfactory results for many large supervised/unsupervised machine learning challenges [57]. In addition, they can handle complex linear and non-linear problems [58]. ANNs can also achieve more accurate and robust prediction results than linear models [59], and may surpass conventional statistical techniques [59], [60]. Although ANNs have demonstrated their ability to handle remotely sensed data [61], most published studies on remote water quality monitoring used linear regression [2].

ANN comprises a set of connected units called neurons that keep the biological concept of brain neurons [57]. The neurons are spread over many fully connected neural networks. The neurons take input $(X_1, X_2, \ldots.., X_m)$ as the first step, collect the input with their internal state (activation) and an optional threshold using an activation function (the second step as shown in Fig. 4(a), and output the result (output value y) based on an output function. There are connections between neurons, each connection supply the output of one neuron as an input to another neuron [57]. The connections have weights $(W_1, W_2, \ldots.., W_m)$ that indicate their relative importance (see Fig. 4(a)). A given neuron can have multiple input connections. The input of a neuron is calculated from the output of their previous neurons and their weighted sum by applying the propagation function. A bias term is usually added to the propagation result (see Fig. 4(a)). The neurons are distributed over multiple layers, one input layer, hidden layers -at least one- and one output layer (see Fig. 4(b)).

B. WHALE OPTIMIZATION ALGORITHM

WOA is a search algorithm [62] that is based on mimicking the social behavior of humpback whales in their hunting strategy. Humpbacks whales track the location of prey and surround them. The optimal design position is initially unknown; thus, the WOA assumes that the target prey is the current best candidate solution or is close to the optimum. Finally, the optimal solution becomes known, and all search space agents update their positions with respect to the optimal value. Equations (1) and (2) illustrate this behavior as follows:

$$\overrightarrow{W} = \mid \overrightarrow{C} . \overrightarrow{X}^*(t) - \overrightarrow{X}(t) \mid \quad (1)$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X}^*(t) - \overrightarrow{A} . \overrightarrow{W} \quad (2)$$

Here, the current iteration is denoted t, $\overrightarrow{W}$ refers to the distance of the ith whale to the best solution obtained so far, $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors, and $\overrightarrow{X}^*$ represents the best solution position vector for the current iteration t. In each iteration, $\overrightarrow{X}^*$ should be updated based on the results of the better solution. The coefficient vectors $\overrightarrow{A}$ and $\overrightarrow{C}$ are measured as illustrated in equations (3) and (4).

$$\overrightarrow{A} = 2\overrightarrow{a}\,\overrightarrow{r} - \overrightarrow{a} \quad (3)$$

$$\overrightarrow{C} = 2\overrightarrow{r} \quad (4)$$

To balance exploration and exploitation, vector $\overrightarrow{a}$ is used. It is linearly reduced over the iterations from *2 to 0*. $\overrightarrow{r}$ is a random vector ranging from *0 to 1* to ensure that all agents can wander in any position in the search space, thus guaranteeing exploration. Equation (2) allows any agent in the search space to update its position according to the current best solution. At the same time, this equation achieves a simulation of encircling the prey. Another behavior of humpback whales in hunting prey called the bubble-net strategy represents the exploitation phase in the algorithm. The two following approaches mathematically model this behavior:

- Shrinking encircling mechanism: The value of $\overrightarrow{A}$ should decrease by reducing the $\overrightarrow{a}$ value in equation (3). From another perspective, $\overrightarrow{A}$ is a random value ranging from $-a$ to a, where a has a value decreasing from *2 to 0* over the iterations. The agent in the search space updates its value with a new value anywhere between the original position and the position of the current best agent.

- Spiral updating position: This technique first measures the distance between the whale position (agent position in the search space) and the prey position (current best solution). Then, a spiral equation (5) is modeled between both the whale and prey positions to imitate humpback whales in their helix-shaped movement.

$$\overrightarrow{X}(t+1) = \overrightarrow{W'}.e^{bl}.cos(2\pi l) + \overrightarrow{X^*}(t) \tag{5}$$

where the distance between the whale and its prey is $\overrightarrow{W}$, b represents the shape of the spiral, and l denotes a random value with the limited range [-1, 1].

Humpback whales on a route to hunt their prey may take either a shrinking circle or spiral-shaped route. For the algorithm to exhibit biomimicry and be closer to nature, a probability of 50% is imposed to select the shrinking encircling mechanism or spiral model in updating the position of the whale. This can be mathematically modeled as in equation (6):

$$\overrightarrow{X^*}(t+1) = \begin{cases} \overrightarrow{X^*}(t) - \overrightarrow{A}.\overrightarrow{W}, & if\ K < 0.5 \\ \overrightarrow{W'}.e^{bl}.cos(2\pi l) + \overrightarrow{X^*}(t), & if\ K \geq 0.5 \end{cases} \tag{6}$$

To ensure the exploitation of the algorithm, $\overrightarrow{A}$ is used with random values less than -1. If $\overrightarrow{A}$ has a value greater than 1, the search agent moves far away from the reference whale, which guarantees exploration. Then, the search agent position is updated according to a randomly selected search agent instead of the best search agent of the current iteration. Thus, $\overrightarrow{A} > 1$ ensures global search. Equations (7) and (8) present a mathematical model of this exploitation process.

$$\overrightarrow{W} = \mid \overrightarrow{C}.\overrightarrow{X_{rand}} - \overrightarrow{X} \mid \tag{7}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X_{rand}} - \overrightarrow{A}.\overrightarrow{W} \tag{8}$$

Here, $\overrightarrow{X_{rand}}$ is a random whale that is selected from the current population. Generally, the WOA has an effective exploration process that allows the optimization algorithm to avoid falling into local minima [62]. The WOA balances the two main processes of the algorithm (exploration and exploitation), which leads the algorithm to reaches the optimal solution with an adequate convergence time [62], [63].

V. PROPOSED HYBRID WATER QUALITY PREDICTION MODEL

The proposed model consists of the following two phases. 1) *Pre-processing phase*: In this phase, images are resampled and their reflectance values are extracted. 2) *Binary WOA (BWOA) and ANN (BWOA-ANN) prediction phase*: In this phase, the BWOA-ANN is proposed to calculate the water quality parameter values. Figure 5 illustrates the BWOA-ANN process.

A. PRE-PROCESSING PHASE

This phase aims to atmospherically correct Sentinel-2A L1C images, and resample the Sentinel-2A L1C bands using SNAP to equalize the spatial resolution of all bands and simplify the handling of different bands of the image. Then, the Bottom of the Atmosphere (BOA) reflectance values are extracted for specific pixels that belong to the collected samples, as illustrated in Algorithm (1).

Algorithm 1 Pre-Processing

1: **Input**: Sentinel-2A images.
2: Using Sentinel-2A toolbox of SNAP which contains C2RCC atmospheric correction to atmospherically correct the images.
3: Resample the bands of each image to 10 m.
4: Extract the reflectance values of the previously mentioned in situ samples (with the information of the coordinates) from all bands of Sentinel-2A images.
5: **Output**: 13 vectors (bands) with BOA reflectance values of samples coordinates $V_s = (RB_1, RB_2, \ldots..RB_{13})$.

B. BIO-INSPIRED HYBRID BWOA-ANN PREDICTION PHASE

There are two main challenges in this phase. First, existing models cannot be applied for different water sources, seasons, or remote sensor data, especially with hyper-spectral space-borne sensors and airborne sensors [2], [64]. This problem exists due to the instability of water from one source to another [2], the change in the spectral and spatial resolution of bands from one sensor to another, and the difference in water quality parameter concentration in different seasons [12]. Therefore, the BWOA is integrated with the ANN to select or optimize the influencing features (bands) and avoid selecting constant bands whose effect varies for different sensors or seasons [2]. Finally, the model predicts the values of water quality parameters. Algorithm (2) presents the detailed steps of this phase. The BWOA has a binary search space that takes the value 0 or 1. Each search agent is represented in an n-dimensional space. For instance, a search agent in a 4-dimensional search space (e.g., 1010) acts as the solution where the first and third features are selected. The Root Mean Square Error (RMSE), as expressed in Equation (9), is used as a fitness function that is minimized to obtain the optimal features.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(S_i - E_i)^2} \tag{9}$$

where E_i represents the observed values, S_i represents the estimated values, and n represents the number of observations. The second challenge in the prediction phase is that the existing studies do not thoroughly examine non-optical parameters, which is due to the poor optical features of these parameters [2], [23], [24]. Therefore, the proposed algorithm is applied to the optical parameters, and after ensuring high R^2 value with minimum RMSE value, the predicted optical parameters are used as features (in addition to bands). This mechanism improves the non-optical prediction accuracy.

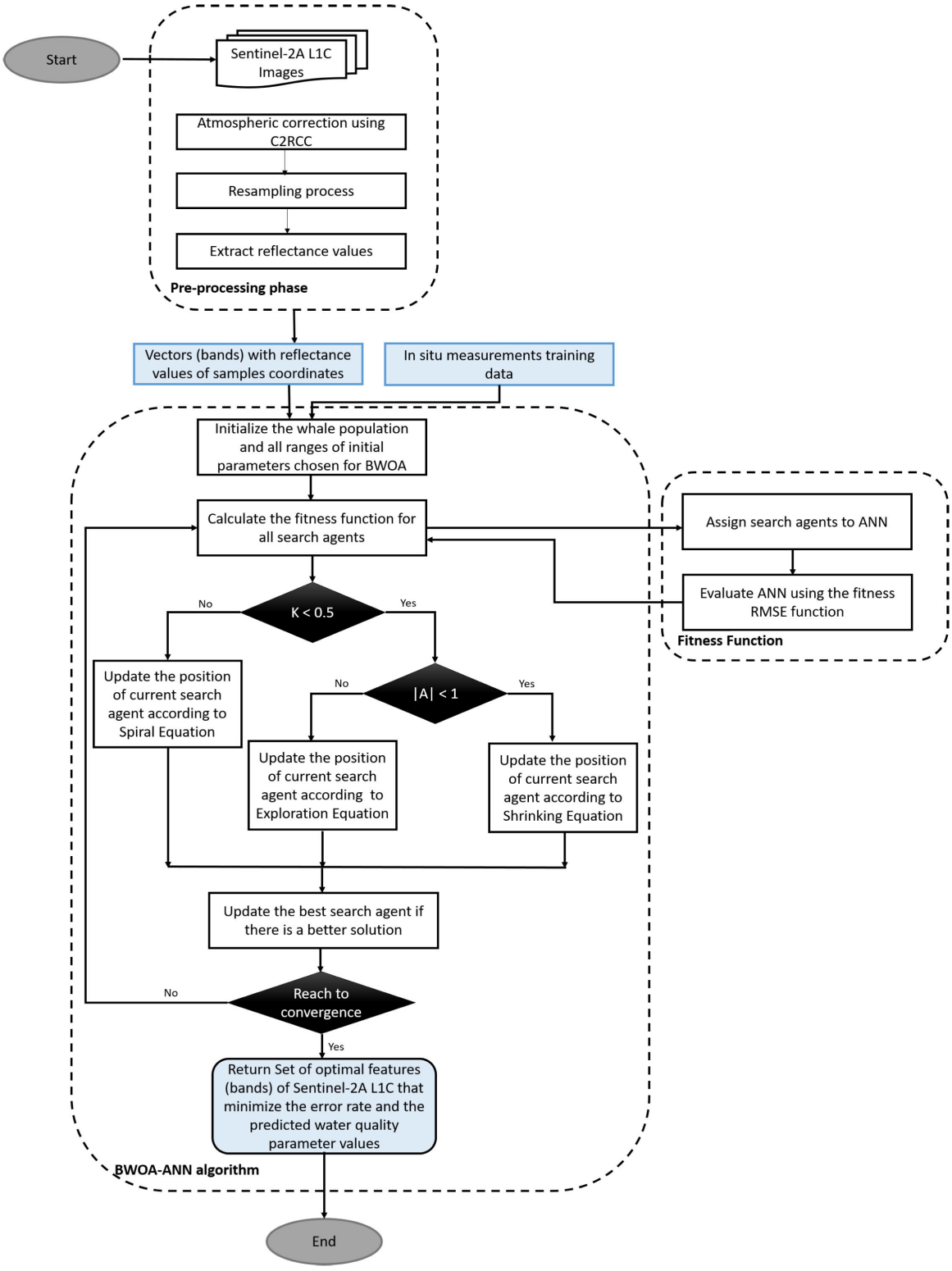

FIGURE 5. Proposed Binary Whale Optimization Algorithm and Artificial Neural Network (BWOA-ANN) model.

VI. EXPERIMENTAL RESULTS AND DISCUSSION

A. IN SITU DATA

Table 2 presents the laboratory measurements of the collected samples of dataset-1 (Nasser Lake, summer [August] 2016) in terms of 12 water quality parameters. Dataset-1 has five sectors from south to north as follows: Abu-Simbel, Tushka, El Madiq, Dahmit, and Aswan. For each sector, three sites were selected. The code section presents the site located

Algorithm 2 Proposed BWOA-ANN

1: **Input**: V_s, all features from the previous phase, and in situ measurements training data.
2: Begin: Generate search space solutions, where the values of whales are randomly generated {0,1}, and each whale is represented as n-dimensional (n value is assigned according to the feature number).
3: Initially, for each search agent in the search space, the fitness value f is calculated according to Equation (9), and f_s, the features selected in a search agent, are calculated.
4: Randomly select X^* (the best search agent) as the initial value.
5: Verify whether the error rate becomes stable.
6: Update a, A, C, l, and K values.
7: Verify whether K is less than 0.5.
8: If the condition in step 7 is satisfied check again if $| A |$ whether less than 1.
9: If the condition in step 8 is satisfied, the position of the current search agent X_c is updated by Equation (1).
10: If the condition in step 8 is not satisfied ($| A |$ is greater than or equal to 1), X_{rand} (a random search agent) is selected, and the position of the current search agent is updated according to Equation (8).
11: If the condition in step 7 is not satisfied (K is greater than or equal to 0.5), the position of the current search agent is updated according to Equation (5).
12: Repeat step 6 for all search agents in the search space.
13: Perform ANN on each whale in the population and calculate f, the fitness value according to Equation (9), on the training set.
14: If there is a better solution, X^* is updated, and if $f(X_c)$ is equal to $f(X^*)$ and $f_s(X_c) < f_s(X^*)$, then update X^* with X_c.
15: Repeat step 5 as long as the condition is satisfied.
16: **Output**: Set of optimal features (bands) of Sentinel-2A L1C that minimize the error rate and the predicted water quality parameter values.

inside the sector: eastern (E), western (W), middle of the main channel (M), and surface water (S). Dataset-2 (Nasser Lake, spring [April] 2016) contained 11 sectors from south to north as follows: Arkeen, Sara, Adendan, Abu-Simbel, Tushka, Ebreem, Wadi El Arab, El Madiq, El Alaaki, Gurf Hussein, and Kalabsha. The samples in dataset-2 were collected with the same procedure as for dataset-1. For each sector, three samples were selected with ES, MS, WS codes. Table 3 presents the laboratory measurements of the collected samples of dataset-2 concerning seven water quality parameters. Table 4 presents the laboratory measurements of 19 samples of dataset-3 (Bin El Ouidane Reservoir, spring [May] 2017) concerning five water quality parameters [24]. Every dataset was divided into a training set and test set at a ratio of 60% and 40%, respectively.

B. SELECTION OF SENTINEL-2 IMAGES

Sentinel-2 images were selected based on several criteria. The selected images were cloud-free and were obtained within at most four days of the in situ measurements [65], [66]. Table 5 presents the Sentinel-2 image dates corresponding to the datasets and water quality samples.

C. PERFORMANCE MEASURES

The accuracy measures of the model were tested using the RMSE, R-squared (R^2), and adjusted R-squared (R^2_{adj}). Additional details of the RMSE are provided in equation (9) in Section (V). R^2 is a factual proportion of how closely the data are fitted to the regression line, and is sometimes called the coefficient of determination. Equation (10) expresses how to calculate R^2, which is the proportion of explained variation $V_{Explained}$ divided by the total variation V_{Total}. R^2_{adj} differs from R^2 that the first consider only the effected independent variables that well interpret the dependent variable. R^2_{adj} can be mathematically calculated in terms of R^2, the number of predictors (independent variables) K, and total sample size N according to equation (11).

$$R^2 = \frac{V_{Explained}}{V_{Total}} \tag{10}$$

$$R^2_{adj} = 1 - \frac{1-N}{N-(K+1)}(1-R^2) \tag{11}$$

Also, the significance value (p-value) was calculated. The p-value measure is commonly used to demonstrate the significant similarity between the estimated and observed results.

D. ATMOSPHERIC CORRECTION RESULTS

Based on the recommendations of the previous studies [47], [67], we applied C2RCC for the atmospheric correction phase to the three datasets. To validate the results of the atmospheric correction, C2RCC retrievals were compared with in situ reflectances measured in the lakes and reservoir. The results of C2RCC are displayed in Table 6, which presents the R^2 and RMSE values concerning Sentinel-2A spectral bands. It can be observed that C2RCC performed acceptable results with R^2 range between 0.61 to 0.98, except for dataset-3 with Band8b is equal to 0.35. RMSE is between 0.0002 to 0.0020.

E. EXPERIMENTAL SCENARIOS

The proposed model was evaluated under different scenarios. The first scenario used only the ANN, while the second scenario used the proposed integrated BWOA-ANN model. The BWOA-ANN was designed to evaluate the best features (bands) for each optical and non-optical water quality parameter and subsequently estimate the water quality parameters. All scenarios were implemented using Python (Spyder 3.3.6) with the Anaconda environment 1.9.7 and tested on a computer with Core i7-8750H, processor (2.20 GHz) and 8 GB of memory.

TABLE 2. Laboratory measurements of the collected samples for the Nasser Lake summer (August) dataset 2016.

Site	Sector	Code	Depth	Chl-a μg/l	Temp 0C	Trans cm	TDS mg/l	TSS mg/l	DO mg/l	COD mg/l	BOD mg/l	PO_4 μg/l	NO_3 μg/l	TN μg/l	TP μg/l
S01	Aswan HD	ES	14.0	10.810	29.500	360.000	162.045	4.120	8.400	8.130	3.900	12.100	23.760	267.012	110.882
S02	Aswan HD	MS	106.0	2.590	28.900	420.000	160.160	5.110	7.360	7.570	2.800	9.900	20.160	253.609	76.165
S03	Aswan HD	WS	23.0	11.330	28.700	500.000	162.760	3.540	7.180	7.410	1.800	11.000	30.960	260.957	78.291
S04	Wadi Abyad	ES	13.0	1.300	30.000	300.000	156.325	4.660	7.430	7.900	2.800	6.600	35.280	213.520	81.302
S05	Wadi Abyad	MS	79.0	0.900	30.200	410.000	155.415	6.020	7.490	7.850	2.900	3.300	15.120	281.247	77.582
S06	Wadi Abyad	WS	32.0	0.900	30.900	460.000	154.115	4.880	7.160	8.280	2.700	6.600	20.880	259.191	75.988
S07	El Madiq	ES	9.0	2.200	30.00	230.00	153.270	6.120	7.640	7.240	2.100	9.900	23.040	322.386	73.508
S08	El Madiq	MS	73.0	2.400	30.300	160.000	154.895	4.120	7.760	7.500	2.600	5.500	27.360	229.364	71.914
S09	El Madiq	WS	26.0	2.400	30.50	220.00	156.000	4.120	7.520	7.840	2.200	8.800	34.560	279.195	72.091
S10	Tushka	ES	16.5	0.840	31.20	120.00	158.405	6.910	8.240	8.200	3.140	11.000	48.960	229.091	68.726
S11	Tushka	MS	60.0	1.180	30.200	130.000	160.095	7.120	8.100	7.990	2.660	11.000	81.260	293.460	83.782
S12	Tushka	WS	16.0	0.790	29.60	90.00	160.875	6.120	8.380	8.460	2.640	9.160	105.120	372.132	69.257
S13	Abu Simble	ES	18.5	10.100	31.00	60.00	166.725	7.160	7.900	8.400	3.020	12.040	41.690	316.280	68.017
S14	Abu Simble	MS	57.5	11.400	30.900	60.000	164.190	8.220	7.800	8.060	2.660	11.290	86.110	316.070	92.107
S15	Abu Simble	WS	39.0	10.300	30.50	80.00	161.850	8.690	8.000	8.330	2.440	36.300	72.000	320.679	93.169

TABLE 3. Laboratory measurements of the collected samples for the Nasser Lake spring (April) dataset 2016.

Site	Sector	Code	Depth	Chl-a μg/l	Temp 0C	Turbidity NTU	TSS mg/l	DO mg/l	TN μg/l	TP μg/l
S01	Kalabsha	ES	30	5.5	26.6	4.20	6	8.24	250.457	90.46
S02	Kalabsha	MS	30	10.2	22.9	3.30	5	6.42	257.28	80.27
S03	Kalabsha	WS	30	4.3	21.6	2.70	4	4.38	241.59	91.27
S04	Gurf Hussein	ES	42	3.2	26.5	4.90	6	7.33	217.37	76.11
S05	Gurf Hussein	MS	42	8.6	22.7	1.70	3	4.70	279.56	80.14
S06	Gurf Hussein	WS	42	8.2	21.7	1.60	2	4.35	212.42	72.64
S07	El Alaaki	ES	55	2.5	27.2	1.70	2	8.03	230.28	64.31
S08	El Alaaki	MS	55	2.5	20.3	2.60	3	4.67	226.46	72.92
S09	El Alaaki	WS	55	2.5	19.7	1.40	2	4.65	236.18	75.62
S10	El Madiq	ES	75	2.4	27.2	1.70	2	8.03	343.44	61.57
S11	El Madiq	MS	75	2	19.8	2.1	3	3.89	312.37	97.04
S12	El Madiq	WS	75	2.4	19.0	1.5	3	3.83	367.28	84.83
S13	Wadi El Arab	ES	55	2	29.0	2.8	4	7.6	225.69	73.74
S14	Wadi El Arab	MS	55	3.7	23.2	1.9	3	4.35	237.75	83.16
S15	Wadi El Arab	WS	55	2.2	20.5	2.3	3	4.07	250.66	79.46
S16	Ebreem	ES	40	7.1	29.1	2.1	3	7.5	335.94	78.64
S17	Ebreem	MS	40	7.3	25.2	2.9	4	4.59	342.88	98.18
S18	Ebreem	WS	40	7	24.2	2.1	4	4.14	304.84	89.77
S19	Tushka	ES	40	4.8	28.0	1.90	4	7.55	360	66.80
S20	Tushka	MS	40	4.6	27.1	2.10	4	7.20	347.64	90.37
S21	Tushka	WS	40	3.2	24.7	2.80	4	5.10	295.71	67.39
S22	Abu Simble	ES	52	7	27.3	2.30	4	8.27	373.26	96.42
S23	Abu Simble	MS	52	7.2	25.3	3.00	5	5.50	334.06	94.75
S24	Abu Simble	WS	52	7.1	20.5	2.50	3	3.80	298.58	86.73
S25	Adendan	ES	43	8.5	28.0	4.00	5	6.78	284.66	79.69
S26	Adendan	MS	43	8.5	26.6	4.50	6	6.64	305.66	92.61
S27	Adendan	WS	43	8.1	26.9	4.70	5	6.52	316.67	81.02
S28	Sara	ES	30	25	26.6	9.50	12	6.78	372.19	96.46
S29	Sara	MS	30	27	26.7	6.90	9	6.50	283.71	76.75
S30	Sara	WS	30	25	26.5	8.00	10	6.40	335.49	84.35
S31	Arkeen	ES	30	32	28.1	9.50	13	6.94	243.36	87.65
S32	Arkeen	MS	30	30	27.4	6.90	11	6.72	301.49	79.19
S33	Arkeen	WS	30	29.8	27.0	8.00	7	6.40	298.63	84.37

In an ANN, the number of neurons in the input layer is equal to the input number, while the number of neurons in the output layer is equal to the number of outputs [56]. The number of neurons in the hidden layer influences both forecasting precision and the convergence of the model. Various operators can affect the number of neurons in the hidden layer, such

TABLE 4. Laboratory measurements of the Bin El Ouidane Reservoir spring (May) dataset 2017 [24].

Site	Chl-a μg/l	DO mg/l	PO_4 mg/l	NO_3 μg/l	TP μg/l
S01	1.89	7.01	0.11	11.21	0.01
S02	0.54	7.01	0.12	10.99	0.01
S03	1.12	7.01	0.12	10.79	0.01
S04	1.00	6.97	0.07	10.14	0.01
S05	1.00	7.38	0.09	12.48	0.01
S06	1.65	7.45	0.08	9.94	0.01
S07	1.73	7.31	0.03	11.29	0.02
S08	1.77	7.56	0.44	10.73	0.01
S09	1.72	7.49	0.05	11.51	0.01
S10	1.85	7.49	0.06	11.14	0.01
S11	1.86	7.56	0.46	11.04	0.01
S12	1.89	7.27	0.05	13.23	0.01
S13	0.54	7.75	2.04	13.16	0.01
S14	1.03	7.82	0.07	13.25	0.01
S15	1.00	8.12	0.07	11.86	0.01
S16	1.00	7.93	0.15	11.58	0.01
S17	1.65	7.97	1.39	11.88	0.01
S18	1.22	7.71	5.45	11.68	0.01
S19	1.00	7.31	0.20	11.90	0.01

as the number of hidden layers, the dataset dimensions, and the number of computations of the activation function [58]. Many studies have presented a variety of rules that specify the number of hidden neurons [58], and the network that is effective on the test set with a minimal number of hidden neurons should be selected [58].

In the proposed model, the neural network was modeled in Keras with three layers (one input layer, one hidden layer, and one output layer). The number of input layer neurons was changed according to the optimal features selected by the BWOA, while the output layer consisted of one output variable (the target water quality parameter). The number of neurons in the hidden layer was set to 8. The rectified linear unit (ReLu) activation function was used for neurons in the hidden layer, while the linear activation function was used for neurons in the output layer.

1) TRADITIONAL ANN EVALUATION

This scenario aimed to test the prediction accuracy of applying only the ANN to the extracted reflectance values from the pre-processing phase for all 13 bands to estimate the optical and non-optical water quality parameters. The results of the ANN are displayed in Table 7, which presents the R^2 and RMSE values for each optical and non-optical parameter. The ANN generated R^2 values for all optical parameters with a mean of 0.51, 0.51, and 0.54 for the test data of dataset-1, dataset-2, and dataset-3, respectively. For all non-optical parameters, the ANN achieved R^2 values with a mean of 0.36, 0.39, and 0.34 for the test data of dataset-1, dataset-2, and dataset-3, respectively. These results are unsatisfactory, however, and can be improved.

2) BWOA-ANN EVALUATION

The proposed model, BWOA-ANN, was evaluated for two sub-scenarios. The first sub-scenario tested the BWOA-ANN for optical parameter estimation, where each parameter was estimated separately. The second sub-scenario tested the model for non-optical parameter estimation using the results from the first sub-scenario. The BWOA is a parameterized algorithm; therefore, the parameters must first be assigned with appropriate values [62] to ensure that the algorithm performs with fast convergence. Table 8 presents the initial parameters of the BWOA algorithm.

In the first sub-scenario, the BWOA-ANN was used to select the appropriate bands for only the optical parameters and then estimate the water quality optical parameters. Table 9 reports the experimental results of this sub-scenario for the three datasets. In addition, it presents the Sentinel-2A bands' relationship with each optical parameter, and illustrates the R^2 and RMSE values for each parameter. Table 9 indicates that instead of performing many manual trials to determine the optimal Sentinel-2A bands for each parameter, the BWOA-ANN can automatically identify the bands as follows. Three common features over the three datasets (Band3, Band5, and Band6) were automatically selected for Chl-a estimation. Previous studies specified that a reflectance peak height between 700 and 720 nm should be used for Chl-a estimation [2]. Three common features over the datasets (Band3, Band4, and Band8a) were the appropriate bands of Sentinel-2A for TDS estimation. This ensured that our proposed model achieved accurate results, as dissolved solids absorption is the highest in the blue band and diminishes exponentially with an increasing wavelength [2]. The BWOA-ANN generated R^2 values over all optical parameters with a mean of 0.93, 0.92, and 0.90 for dataset-1, dataset-2, and dataset-3, respectively, when the model was applied to the test dataset. As illustrated in Table 9, the BWOA-ANN proved to be robust against different optical parameters over different datasets.

A stopping criteria was set such that the algorithm stopped when it reached the convergence (stable) state (i.e., best optimal features), and this occurred when the proposed model detected that there were no more changes in the RMSE value (fitness function). In this scenario, the maximum number of BWOA iterations to reach the optimal value (best solution) was 16 with an average central processing unit (CPU) time of 36.5661 s for one parameter.

Also, the BWOA-ANN was used to select the appropriate bands for the non-optical parameters and estimates them. The BWOA-ANN generated R^2 values over all non-optical parameters with a mean of 0.61, 0.53, and 0.60 for dataset-1, dataset-2, and dataset-3, respectively, when the model was applied to the test dataset. But these results could be improved. After ensuring that the optical parameters were predicted with an R^2 value greater than 0.90 in the first sub-scenario and based on the conclusion that there are relationships between the different water quality parameters [68], these optical parameters were added as features in addi-

TABLE 5. Dates of Sentinel-2A L1C satellite images and respective corresponding datasets and water quality samples.

Sentinel-2A L1C image date	Dataset	Number of samples	Water quality sampling date
August 04, 2016	Nasser Lake summer (August)	6 samples of Aswan HD and Wadi Abyad sectors	August 05, 2016
August 04, 2016	Nasser Lake summer (August)	6 samples of El-Madiq and Tushka sectors	August 08, 2016
August 14, 2016	Nasser Lake summer (August)	3 samples of Abu-Simbel sectors	August 15, 2016
April 15, 2016	Nasser Lake spring (April)	3 samples of Kalabsha sector	April 11, 2016
April 15, 2016	Nasser Lake spring (April)	6 samples of Gurf Hussein and El Alaaki sectors	April 14, 2016
April 15, 2016	Nasser Lake spring (April)	3 samples of Al Madeek sector	April 15, 2016
April 15, 2016	Nasser Lake spring (April)	3 samples of Wadi El Arab sector	April 16, 2016
April 15, 2016	Nasser Lake spring (April)	3 samples of Ebreem sector	April 17, 2016
April 15, 2016	Nasser Lake spring (April)	3 samples of Toushka sector	April 19, 2016
April 25, 2016	Nasser Lake spring (April)	3 samples of Abu Simble sector	April 21, 2016
April 25, 2016	Nasser Lake spring (April)	3 samples of Arkeen sector	April 23, 2016
April 25, 2016	Nasser Lake spring (April)	6 samples of Adendan and Sara sectors	April 25, 2016
May 26, 2017	Bin El Ouidane Reservoir (May)	19 samples from the downstream part of the Bin El Ouidane Reservoir	May 26, 2017

TABLE 6. Performance measurements of the atmospheric correction phase.

	Band number	R^2	RMSE
Dataset-1 Nasser Lake summer (August)	Band1	0.7062	0.0020
	Band2	0.7558	0.0015
	Band3	0.9842	0.0016
	Band4	0.9803	0.0006
	Band5	0.9773	0.0005
	Band6	0.9351	0.0002
	Band7	0.9491	0.0002
	Band8b	0.7842	0.0015
Dataset-2 Nasser Lake spring (April)	Band1	0.6927	0.0014
	Band2	0.8038	0.0016
	Band3	0.9596	0.0013
	Band4	0.9740	0.0006
	Band5	0.9721	0.0003
	Band6	0.9363	0.0002
	Band7	0.9028	0.0003
	Band8b	0.7537	0.0014
Dataset-3 Bin El Quidane Randeservoir (May)	Band1	0.5690	0.0018
	Band2	0.6105	0.0002
	Band3	0.8941	0.0016
	Band4	0.8727	0.0008
	Band5	0.8961	0.0008
	Band6	0.8207	0.0012
	Band7	0.8061	0.0001
	Band8b	0.3533	0.0002

TABLE 7. Optical and non-optical parameter performance measurements for traditional Artificial Neural Network (ANN).

Optical activity	Parameter	Dataset-1 Nasser Lake summer (August) 2016				Dataset-2 Nasser Lake spring (April) 2016				Dataset-3 Bin El Ouidane Reservoir (May) 2017			
		Training		Testing		Training		Testing		Training		Testing	
		R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Optical parameter	Chl-a	0.6149	11.31768	0.5584	20.35729	0.5706	7.27574	0.5237	9.31957	0.5846	4.24753	0.5408	7.27519
	TDS	0.5849	9.24374	0.5462	29.31158	-	-	-	-	-	-		-
	TSS	0.5079	5.3497	0.4479	14.3845	0.5770	4.28497	0.4972	7.3085	-	-	-	-
	Trans	0.5342	4.21785	0.4604	8.25706	-	-	-	-	-	-	-	-
	Turbidity	-	-	-	-	0.6182	1.9724	0.5846	5.2897	-	-	-	-
Non-optical parameter	DO	0.3877	22.4478	0.3507	30.4949	0.3304	6.40547	0.30278	9.44570	0.3975	10.47520	0.3574	30.45924
	COD	0.3782	4.48675	0.3038	9.57366	-	-	-	-	-	-	-	-
	BOD	0.4397	1.27356	0.4094	7.46918	-	-	-	-	-	-	-	-
	PO_4	0.3951	11.8694	0.3515	20.6789	-	-	-	-	0.3720	8.6942	0.3158	14.7195
	NO_3	0.3666	9.5343	0.3264	23.60380	-	-	-	-	0.4094	9.7206	0.3943	17.7490
	TN	0.3822	13.43587	0.3448	40.5490	0.5161	13.3428	0.4957	20.2467	-	-	-	-
	TP	0.4267	10.4571	0.3678	18.4861	0.4916	12.4280	0.3924	21.1286	0.3785	4.6842	0.3472	6.7306

TABLE 8. Initial parameters selected for Binary Whale Optimization Algorithm (BWOA) implementation.

Parameter	Value
Maximum number of iterations	100
Population size	50
Range of search agent	{0,1}
Vector a linearly decreased in range	[2,0]
Problem dimensions	No. of features in the data

tion to the Sentinel-2A bands in the second sub-scenario to enhance the detection of non-optical parameters. Table 10 presents the experimental results of the second sub-scenario. This table presents the relationship between the Sentinel-2A bands and each non-optical parameter. In addition, it illustrates the correlated optical parameters, and tabulates the R^2 and RMSE values for each parameter. These results were obtained with a significant correlation between the different chemical parameters in several water sources [68]. The BWOA-ANN generated R^2 values over all non-optical parameters with a mean of 0.89, 0.89, and 0.87 for dataset-1, dataset-2, and dataset-3, respectively, when the model was applied to the test dataset. The R^2 and RMSE values were preserved with few changes for different non-optical parameters. However, the problem dimension increased, as optical parameters other than bands were introduced as additional features. Nonetheless, the model accuracy was not negatively affected, which demonstrates the robustness of the proposed model against various water quality parameters and increased problem dimensions. In this experiment, the maximum number of BWOA iterations to reach the optimal value (best solution) was 24 with an average CPU time of 59.792 s for one parameter.

Also, Table 11 presents the average R^2, R^2_{adj}, and p-value measurements of the proposed BWOA-ANN prediction model considering the three datasets over 30 independent runs. It was observed that the p-value was less than 4.35E-03 over the three datasets, which indicated the high-level significance relation between the estimated and observed results. Moreover, the generated R^2 values over all optical parameters with a mean of 0.93, 0.92, and 0.88 with R^2_{adj} values with a mean of 0.92, 0.92, and 0.87 for dataset-1, dataset-2, and dataset-3, respectively, when the model was applied to the test dataset. For non-optical parameters, the generated R^2 values with a mean of 0.90, 0.89, and 0.88 with R^2_{adj} values with a mean of 0.88, 0.87, and 0.87 for dataset-1, dataset-2, and dataset-3, respectively, when the model was applied to the test dataset.

Both ANN and BWOA-ANN models were tested against optical and non-optical parameters over the datasets. Table 7, 9, and 10 reveal that the results obtained from the BWOA-ANN model were superior to those from the traditional ANN, which doesn't apply the band selection process, with the R^2 value approximately 40% higher for optical parameters and up to 52% higher for non-optical parameters. Generally, the BWOA-ANN demonstrated vast improvement over the traditional ANN with good generalization over datasets from different locations with diverse conditions. The proposed model can thus be applied to a range of optical and non-optical water quality parameter estimation scenarios while preserving its generalization, as illustrated in Table 9 and 10.

Our model can update itself according to the water source or season due to the automatic feature selection process. This process selects the appropriate bands and features for the season or location. The model does not require modifications when applying it to another place or another season, only by training it for the first time on a specific lake, thus it can be used to identify the water quality parameters for the same lake in any other season. Generally, to prove that the model has a good generalization performance, we test the model on four different seasons for Nasser lake after the model had been trained on dataset-1. Fifteen water samples were collected from the lake for each season with the same method of dataset-1 and dataset-2. Table 12 presents the average R^2, R^2_{adj}, and RMSE measurements of the proposed BWOA-ANN prediction model considering the four seasons of Nasser lake. It was observed that the generated R^2 values over all available optical parameters with a mean of 0.91, 0.92, 0.94, and 0.95 with R^2_{adj} values with a mean of 0.89, 0.90, 0.90, and 0.91 and RMSE values with a mean of 3.32, 3.30, 1.76, and 2.74 for seasons (March) 2019, (July) 2019, (November) 2019, and (June) 2020, respectively. For non-optical parameters, the generated R^2 values with a mean of 0.89, 0.88, 0.92, and 0.90 with R^2_{adj} values with a mean of 0.88, 0.87, 0.90, and 0.88 and RMSE values with a mean of 3.39, 3.75, 2.41, and 2.83 for seasons (March) 2019, (July) 2019, (November) 2019, and (June) 2020, respectively. Based on the presented results, it can be observed that the proposed model achieved stable results against the different seasons and different optical and non-optical parameters.

The proposed model was subsequently used to clarify the current maps of the year 2020 to evaluate the quantitative water quality estimation model. Two Sentinel-2A L1C scenes were selected to investigate several water quality parameters. The first scene was for Nasser Lake on May 30, 2020, where Chl-a, Turbidity, COD, and DO were identified as parameters to be estimated. The second Sentinel-2A L1C scene was selected for the Bin El Ouidane Reservoir on June 27, 2020, where DO, TP, and NO_3 were identified as parameters to be estimated. The trained BWOA-ANN was applied for each parameter to the selected scenes, generating the maps illustrated in Figure 6. Figure 6(a-d) illustrates the variation of the four parameters for Nasser Lake. The Chl-a concentration varied between 2 and 30 μg/l and was generally less than 4 over the entire lake, which is considered excellent quality according to Egyptian Governmental Decree No. 92/2013 [69]. The far south of the lake had high values of Chl-a ranging between 25 and 30 μg/l, which is

TABLE 9. Optical parameter performance measurements and their corresponding Sentinel-2A bands for the proposed Binary Whale Optimization Algorithm and Artificial Neural Network (BWOA-ANN) model.

Parameter	Dataset-1 Nasser Lake summer (August) 2016					Dataset-2 Nasser Lake spring (April) 2016					Dataset-3 Bin El Ouidane Reservoir (May) 2017				
	Training		Testing		Correlated bands	Training		Testing		Correlated bands	Training		Testing		Correlated bands
	R^2	RMSE	R^2	RMSE		R^2	RMSE	R^2	RMSE		R^2	RMSE	R^2	RMSE	
Chl-a	0.9421	0.8027	0.9372	0.8786	B3, B4, B5, B6	0.9649	0.3348	0.9573	0.8034	B1, B3, B5, B6	0.9126	0.1272	0.8967	0.4768	B3, B5, B6
TDS	0.9403	3.1498	0.9197	5.0549	B3, B4, B8a, B8b	-	-	-	-	B3, B4, B8a, B8b	0.9434	0.7170	0.9146	1.2067	B3, B4, B8a
TSS	0.9548	0.1842	0.9469	1.6578	B3, B4, B8a, B8b	0.9173	0.9164	0.8864	1.1440	B4, B8a, B8b	0.9172	0.9157	0.9038	2.2760	B4,B5, B8a, B8b
Trans	0.9673	3.6267	0.9470	5.1185	B1, B3	-	-	-	-	-	-	-	-	-	
Turbidity	-	-	-	-	-	0.9728	0.8768	0.9346	1.2497	B1, B2, B3	-	-	-	-	-

TABLE 10. Non-optical parameters performance measurements, correlated bands, and their Sentinel-2A correlated bands for the proposed Binary Whale Optimization Algorithm and Artificial Neural Network (BWOA-ANN) model.

Parameter	Dataset-1 Nasser Lake summer (August) 2016						Dataset-2 Nasser Lake spring (April) 2016						Dataset-3 Bin El Ouidane Reservoir (May) 2017					
	Training		Testing		Correlated bands	Optical correlated parameters	Training		Testing		Correlated bands	Optical correlated parameters	Training		Testing		Correlated bands	Optical correlated parameters
	R^2	RMSE	R^2	RMSE			R^2	RMSE	R^2	RMSE			R^2	RMSE	R^2	RMSE		
DO	0.9151	1.0491	0.8949	1.5795	B3, B5, B6	Chl-a	0.9349	0.6167	0.9249	0.9781	B3, B5, B6	Chl-a	0.8540	1.8249	0.8338	1.9392	B5, B8a, B11	Chl-a
COD	0.9248	0.8437	0.8903	1.2533	B3, B4	TDS, Trans	-	-	-	-	-	-	-	-	-	-	-	-
BOD	0.9144	0.5845	0.8972	1.4169	B3, B4, B8a, B8b	Chl-a, Trans	-	-	-	-	-	-	-	-	-	-	-	-
PO_4	0.9403	2.1613	0.9184	2.5080	B1, B3, B4, B5, B8a	Chl-a	-	-	-	-	-	-	0.9544	1.6187	0.9046	3.0930	B1, B3, B8a, B11	Chl-a
NO_3	0.8942	1.1684	0.8896	3.7140	B1, B2, B3	Chl-a, TDS	-	-	-	-	-	-	0.9043	2.7590	0.8946	5.0279	B1,B3, B11	Chl-a
TN	0.9348	10.6782	0.8964	16.1864	B2, B4, B7	Chl-a, TDS, Trans	0.9024	5.1038	0.8831	9.9000	B2, B4, B7	Chl-a	-	-	-	-	-	-
TP	0.9567	3.6472	0.90167	5.6207	B2, B4, B7	Chl-a, TSS, Trans	0.8936	7.1468	0.8804	7.8073	B2, B4, B7, B8	Chl-a	0.9146	0.0218	0.8831	0.0464	B2, B4, B7	Chl-a

TABLE 11. Performance measurements of the proposed BWOA-ANN model with testing datasets over 30 independent runs for optical and non-optical parameters.

Parameter	Dataset-1 Nasser Lake summer (August)			Dataset-2 Nasser Lake spring (April)			Dataset-3 Bin El Ouidane Reservoir (May)		
	R^2	R^2_{adj}	p-value	R^2	R^2_{adj}	p-value	R^2	R^2_{adj}	p-value
Chl-a	0.9305	0.9274	1.6642E-10	0.9434	0.9390	1.7396E-09	0.8812	0.8765	1.5843E-06
TDS	0.9246	0.9042	2.1641E-07	-	-	-	-	-	-
TSS	0.9446	0.9342	1.1528E-06	0.9035	0.9003	8.7515E-17	-	-	-
Trans	0.9591	0.9490	8.4829E-20	-	-	-	-	-	-
Turbidity	-	-	-	0.9573	0.9350	3.6811E-12	-	-	-
DO	0.9075	0.8816	3.2409E-03	0.9168	0.9043	1.7256E-11	0.8497	0.8390	6.0844E-05
COD	0.9304	0.9153	2.2674E-09	-	-	-	-	-	-
BOD	0.9073	0.8926	4.3491E-03	-	-	-	-	-	-
PO_4	0.9123	0.8867	6.1388E-07	-	-	-	0.9345	0.9264	8.6524E-13
NO_3	0.8893	0.8707	5.6184E-18	-	-	-	0.9068	0.8834	7.6498E-16
TN	0.8946	0.8837	1.2647E-14	0.8837	0.8649	9.0691E-03	-	-	-
TP	0.8934	0.8873	6.1276E-10	0.8947	0.8704	1.9345E-08	0.8749	0.8637	1.6470E-12

considered poor quality according to Egyptian Governmental Decree No. 92/2013 [69] (see Fig. 6(a)). Turbidity concentration demonstrated excellent values of less than 5 NTU in the majority of the lake and satisfactory values in the far north and south (see Fig. 6(b)). The COD concentration varied along the lake with values between 5 and 15 mg/l, it is not observed any bad ranges for COD along the lake according to Egyptian Governmental Decree No. 92/2013 [69] (see

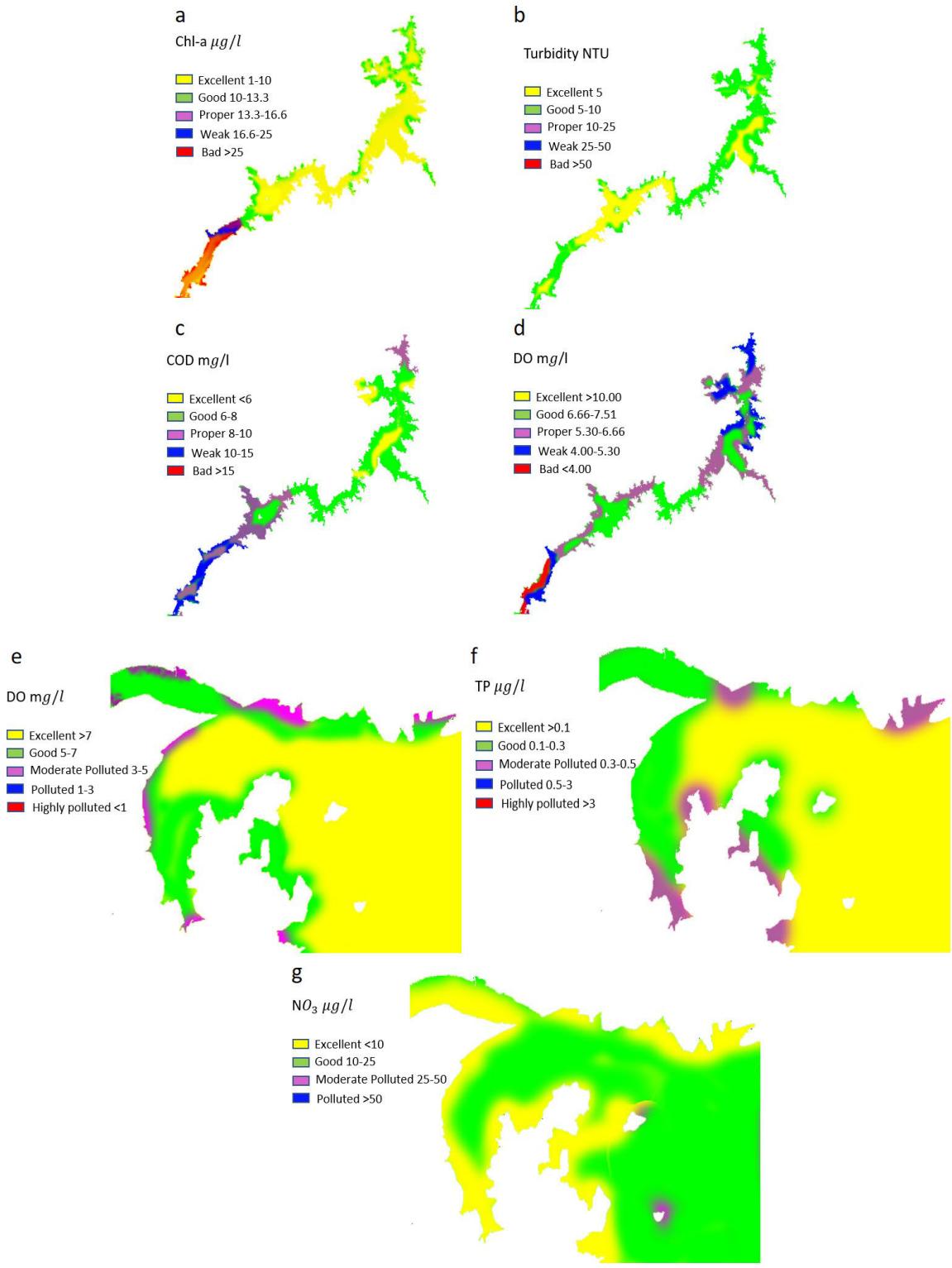

FIGURE 6. Estimated water quality parameters map: a, b, c, and d Nasser lake on May 30,2020; e, f, and g Bin El Ouidan Reservoir on June 27, 2020.

Fig. 6(c)). With respect to the DO concentrations of the lake, good values ranged between 6.66 and 7.51 mg/l in the center of the lake. Proper values were noticed in the range 5.30-6.66 on some lakesides with weak values in the far north and south (see Fig. 6(d)). Figure 6(e-g) presents the second scene maps for the Bin El Ouidane Reservoir, which were

TABLE 12. Performance measurements of the proposed BWOA-ANN model with testing datasets for different seasons of Nasser lake.

Parameter	(March) 2019			(July) 2019			(November) 2019			(June) 2020		
	R^2	R^2_{adj}	RMSE	R^2	R^2_{adj}	RMSE	R^2	R^2_{adj}	RMSE	R^2	R^2_{adj}	RMSE
Chl-a	-	-	-	-	-	-	0.9470	0.9035	0.9349	0.9673	0.9514	1.0428
TDS	0.9149	0.9043	2.2861	0.9408	0.9076	3.1735	0.9317	0.8924	2.1760	0.9649	0.9248	0.7547
TSS	0.9187	0.8830	3.4387	-	-	-	0.9346	0.9043	2.4359	-	-	-
Trans	0.9261	0.8934	4.2640	0.9143	0.9057	3.4267	0.9513	0.9240	1.4384	0.9308	0.8835	6.4237
DO	0.9138	0.8901	3.4297	0.8867	0.8761	5.2468	0.9362	0.9137	1.5430	0.9143	0.8907	3.4299
COD	0.9182	0.9030	3.7288	0.8934	0.8744	5.1275	0.9249	0.8843	2.4085	0.9061	0.8770	6.1725
BOD	0.8850	0.8673	3.4582	0.8675	0.8473	3.4582	0.9137	0.9094	0.5758	0.9037	0.8952	2.2557
PO_4	0.9006	0.8966	3.4387	0.9127	0.9091	1.4672	0.9399	0.9208	0.8346	0.9231	0.8938	1.2647
NO_3	0.8703	0.8680	2.9438	0.8849	0.8810	3.4766	0.8927	0.8873	0.5438	0.8876	0.8652	1.0627
TN	-	-	-	-	-	-	0.9293	0.9143	6.4982	-	-	-
TP	-	-	-	-	-	-	0.9362	0.9043	4.5007	-	-	-

TABLE 13. Comparison of the proposed model and other existing remotely monitor water quality models.

Study	Detected parameters		Automatically band selection (Optimization)	Dynamically optical and non-optical parameters relation investigation	Sensor	Application area	Results
	Optical	Non-optical					
[32] 2015	✓ COD	✗	✗	✗	Landsat-8 Sentinel-2 Sentinel-3	Year: 2013 Samples number: 35 Water source:Mississippi River Location: North America	R^2: 0.84-0.86 R^2_{adj}: 0.85 RMSE: 43.8-43.8
[40] 2016	✓ Chl-a	✗	✗	✗	MODIS	Year: 2013 and 2014 Samples number:249 Water Source: Chaohu Lake, Taihu Lake, Hongzehu Lake, and Poyanghu Lake Location: China	R^2: 0.85 Unbiased RMSE: 57.2 RMSE: 77.9 P-value<0.01
[34] 2017	✓ Chl-a	✗	✗	✗	Sentinel-2	Year: 2016 and 2017 Samples number: 30 Water source: Ba Be Lake Location: Northern Vietnam	R^2: 0.68
[71] 2018	✓ TSS	✓ TN	✗	✗	Landsat-TM, ETM+, and OLI/TIRS	Year: 1983-2015 Samples number: 28 Water source: Lady Bird Lake Location: Texas, USA	**TSS accuracy:** R^2: 0.70 **TN accuracy:** R^2: 0.37
[21] 2018	✓ Chl-a, Secchi disc depth (Trans), and Turbidity	✗	✗	✗	Landsat-8 OLI	Year: 2013 Samples number: 73 Water source: Chivero Lake, and Mazvikadei Lake Location: Zimbabwe	**Chl-a accuracy:** R^2: >0.84 in Mazvikadei Lake R^2: 0.69 in Chivero Lake **Secchi disc depth (Trans) accuracy:** R^2: >70 for both lakes **Turbidity accuracy:** R^2: $\geq$ 0.65 for both lakes
[24] 2019	✓ Chl-a	✓ DO, TP, NO_3, and PO_4	✗	✗	Sentinel-2	Year: 2017 Samples number: 19 Water source: Bin El Ouidane ReservoirLocation: Morocco	R^2: >0.52 RMSE: <1.024
[26] 2019	✓ Chl-a concentration, Turbidity	✓ NO_3	✗	✗	Sentinel-2	Year: 2017-2018 Samples number: 120 Water source: Dam lake of Wadi Baysh Location: Saudi Arabia	**Chl-a accuracy:** R^2: 0.88 R^2_{adj}: 0.884012 RMSE: 0.0732 **Turbidity accuracy:** R^2: 0.92 R^2_{adj}:0.91 RMSE:0.0764 NO_3 **accuracy:** R^2: 0.92 R^2_{adj}:0.91 RMSE: 0.0627 p <0.05
[22] 2020	✓ blue-green algae (BGA), Chl, fluorescent dissolved organic matter (fDOM), specific conductance (SC), and Turbidity	✓ DO	✗	✗	Landsat-8, and Sentinel-2	Year: 2013-2018 Samples number: 97 Water source: La Crosse, Sroddard, Decatur Lake, and Carlyle Lake Location: Midwestern United States	**BGA accuracy:** R^2: 0.91 RMSE: 0.8630 **Chl accuracy**: R^2 0.88 RMSE: 7.561 **DO accuracy:** R^2: 0.89 RMSE: 1.806 **fDOM accuracy:** R^2: 0.92 RMSE: 14.496 **SC accuracy:** R^2: 0.87 RMSE: 48.463 **Turbidity accuracy:** R^2: 0.83 RMSE: 5.190
C2RCC	✓ Chl-a, and TSS	✗	✗	✗	Sentinel-2	Year: 2016, and 2017 Samples number: 67 Water source: Nasser Lake (April 2016) Nasser Lake (August 2016) Bin El Ouidan Reservoir (May 2017) Location: Egypt, Morocco	**Nasser Lake (April 2016)** (Chl-a) R^2: 0.56, R^2_{adj}: 0.36, RMSE: 13.1 (TSS) R^2: 0.30, R^2_{adj}: 0.25, RMSE: 15.2 **Nasser Lake (August 2016)** (Chl-a) R^2: 0.48, R^2_{adj}: 0.42, RMSE: 15.1 (TSS) R^2: 0.37, R^2_{adj}: 0.29, RMSE: 10.2 **Bin El Ouidan Reservoir (May 2017)** (Chl-a) R^2: 0.25, R^2_{adj}: 0.15, RMSE: 18.75
Proposed model	✓ Chl-a, Trans, TDS, Turbidity, and TSS	✓ DO, COD, BOD, PO_4 NO_3, TP, and TN	✓	✓	Sentinel-2	Year: 2016, and 2017 Samples number: 67 Water source: Nasser Lake (April 2016) Nasser Lake (August 2016) Bin El Ouidan Reservoir (May 2017) Location: Egypt, Morocco	**Optical parameters accuracy over three datasets:** R^2: 0.91 R^2_{adj}: 0.90 RMSE: 1.8391 **Non-optical parameters accuracy over three datasets:** R^2: 0.89 R^2_{adj}: 0.886 RMSE: 4.3304 P-value: <4.35E03

ranged according to the water quality levels of Moroccan Official Bulletin [70]. The DO concentration was in the range 5-8 along the water body, and satisfactory values of DO were observed in the center of the reservoir with values greater than 7. The reservoir contained excellent values of TP within the range 0.01-1 in the majority of the reservoir. Concerning the NO_3 concentration, good values ranged between 10 and 25 along the reservoir with excellent values less than 10 along the reservoir sides.

Table 13 illustrates a comparative summary for specifications of the proposed model against several existing remotely monitor water quality models. From Table 13, it is observed that the proposed model outperformed the models proposed in [32], [34], [40], and [21] with R^2 and RMSE ratios. In addition, these proposed models investigates only the optical parameters. Moreover, the proposed model was tested against various water sources, and the model robustness was not negatively affected with all optical and non-optical parameters as can be noticed with the model proposed in [21], which reported instability of Chl-a accuracy with different lakes and weather conditions. Compared to the models proposed in [71], and [24], which investigated some of the optical and non-optical parameters, the proposed model outperforms all of them in terms of R^2 and RMSE. On the other hand, the proposed model shows almost equal accuracy with the proposed models in [26] and [22], but our proposed model was tested against 11 different optical and non-optical parameters and different water sources while preserving the obtained accuracy over three datasets. In addition, the proposed model automatically selects suitable satellite bands for each parameter, and dynamically investigates the relation between optical and non-optical parameters.

Also, we tested the standard C2RCC in Chl-a and TSS estimation with our three datasets. C2RCC has been trained in large scales of scattering and absorption properties, however, it is observed from Table 13 that C2RCC performed little performance comparing with our model. Moreover, the results vary from one season to another for the same lake, with R^2 difference equals to 0.6 and 0.7 for Chl-a and TSS, respectively. On the other hand, it is clear from Table 13 that C2RCC does not work well with small reservoirs (Bin El Ouidan Reservoir), achieving an R^2 value of 0.25 and RMSE value of 18.75. And till now, there is no estimation for non-optical water quality parameters in the C2RCC processor.

Finally, up to our knowledge, the proposed model is the first model, which uses the bands optimization process and dynamically investigates the relation between optical and non-optical parameters in the field of remotely monitor water quality. This optimization process enables the model to be applied to another satellite without needing to change the model steps; thus, the model is applicable to different case studies and satellites.

VII. CONCLUSION AND FUTURE WORK

Many existing studies in remote water quality monitoring have low robustness against changes in geographic regions and seasons, and cannot be applied to other study areas. This paper presents an alternative prediction model that can be applied to various study areas, seasons, and satellite images. The model achieved satisfactory results for both optical and non-optical parameters, and had meaningful relationships between non-optical and optical parameters that were confirmed by previous literature. The results illustrate that Sentinel-2 data have an important role in remote water quality monitoring, as they can perfectly map optical parameters (Chl-a, Trans, TDS, Turbidity, and TSS) and non-optical parameters (DO, COD, BOD, PO_4, NO_3, TP, and TN). The proposed model was tested against three datasets from diverse locations in different seasons obtaining a mean R^2 value of 0.916 for optical parameters and a mean R^2 value of 0.890 for non-optical parameters for three different datasets. Moreover, it was observed that using the band selection process for the proposed BWOA-ANN outperformed the performance of ANN by 40% and 52% for optical parameters and non-optical parameters, respectively. We implemented multiple experiments with three datasets achieving the mean R^2 and R^2_{adj} values of 0.910 and 0.903, respectively, for optical parameters. For non-optical parameters, the model obtained the mean R^2 and R^2_{adj} values of 0.890 and 0.870, respectively. The proposed model outperformed the state-of-the-art in terms of R^2, R^2_{adj}, and RMSE for both optical and non-optical parameters. Also, the proposed model, in general, outperformed the state-of-the-art in terms of automatically selecting the satellite bands and dynamically investigating optical and non-optical parameters relation.

In future work, the BWOA-ANN model can be further applied using other sensors, such as high spectral space-borne sensors and airborne sensors whose spectral bands can reach 224 bands, as the proposed model can automatically select optimal bands. Also, Data fusion can be performed between different sensors to fill temporal gaps.

ACKNOWLEDGMENT

The authors thank Dr. Seliem M. El Sayed and Ayman El Badry, Researchers in NIOF, for helping the chemical field measurements analysis and Dr. Soaad Sabae for providing the data of Chlorophyll-a. They also thank the National Water Research Center, Egyptian Ministry of Water Resources and Irrigation for assisting us with information and chemical data analysis.

REFERENCES

[1] J. A. Le, H. M. El-Askary, M. Allali, E. Sayed, H. Sweliem, T. C. Piechota, and D. C. Struppa, "Characterizing el Niño-southern oscillation effects on the blue Nile yield and the Nile river basin precipitation using empirical mode decomposition," *Earth Syst. Environ.*, vol. 4, no. 4, pp. 699–711, Dec. 2020.

[2] M. Gholizadeh, A. Melesse, and L. Reddi, "A comprehensive review on water quality parameters estimation using remote sensing techniques," *Sensors*, vol. 16, no. 8, p. 1298, Aug. 2016.

[3] G. R. Fones, A. Bakir, J. Gray, L. Mattingley, N. Measham, P. Knight, M. J. Bowes, R. Greenwood, and G. A. Mills, "Using high-frequency phosphorus monitoring for water quality management: A case study of the Upper River Itchen, UK," *Environ. Monitor. Assessment*, vol. 192, no. 3, pp. 1–15, Mar. 2020.

[4] C. Medupin, "Spatial and temporal variation of benthic macroinvertebrate communities along an urban river in greater Manchester, UK," *Environ. Monitor. Assessment*, vol. 192, no. 2, pp. 1–20, Feb. 2020.

[5] L. K. Bhardwaj and T. Jindal, "Persistent organic pollutants in lakes of Grovnes peninsula at Larsemann Hill area, East Antarctica," *Earth Syst. Environ.*, vol. 4, no. 2, pp. 349–358, Jun. 2020.

[6] E. W. Kimani-Murage and A. M. Ngindu, "Quality of water the slum dwellers use: The case of a kenyan slum," *J. Urban Health*, vol. 84, no. 6, pp. 829–838, Nov. 2007.

[7] Z. Sharip, S. B. A. Razak, N. Noordin, and F. M. Yusoff, "Application of an effective microorganism product as a cyanobacterial control and water quality improvement measure in Putrajaya Lake, Malaysia," *Earth Syst. Environ.*, vol. 4, no. 1, pp. 213–223, Mar. 2020.

[8] C. W. Burns, M. Schallenberg, and P. Verburg, "Potential use of classical biomanipulation to improve water quality in New Zealand lakes: A re-evaluation," *New Zealand J. Mar. Freshwater Res.*, vol. 48, no. 1, pp. 127–138, Jan. 2014.

[9] G.-S. Park, A. R. Khan, Y. Kwak, S.-J. Hong, B. Jung, I. Ullah, J.-G. Kim, and J.-H. Shin, "An improved effective microorganism (EM) soil ball-making method for water quality restoration," *Environ. Sci. Pollut. Res.*, vol. 23, no. 2, pp. 1100–1107, Jan. 2016.

[10] M. Lürling, G. Waajen, and L. N. de Senerpont Domis, "Evaluation of several end-of-pipe measures proposed to control cyanobacteria," *Aquatic Ecol.*, vol. 50, no. 3, pp. 499–519, Sep. 2016.

[11] K. Dörnhöfer and N. Oppelt, "Remote sensing for lake research and monitoring—Recent advances," *Ecol. Indicators*, vol. 64, pp. 105–122, May 2016.

[12] L. Andres, K. Boateng, C. Borja-Vega, and E. Thomas, "A review of *in-situ* and remote sensing technologies to monitor water and sanitation interventions," *Water*, vol. 10, no. 6, p. 756, Jun. 2018.

[13] K. Puczko and E. Jekatierynczuk-Rudczyk, "Analysis of urban land cover influence to organic carbon and nutrients in surface water via impacted groundwater," *Environ. Monitor. Assessment*, vol. 192, no. 2, p. 145, Feb. 2020.

[14] C. Z. F. Braga, A. W. Setzer, and L. D. de Lacerda, "Water quality assessment with simultaneous Landsat-5 TM data at Guanabara Bay, Rio de Janeiro, Brazil," *Remote Sens. Environ.*, vol. 45, no. 1, pp. 95–106, Jul. 1993.

[15] H. Abdelmoneim, M. R. Soliman, and H. M. Moghazy, "Evaluation of TRMM 3B42V7 and chirps satellite precipitation products as an input for hydrological model over Eastern nile basin," *Earth Syst. Environ.*, vol. 4, pp. 685–698, Nov. 2020.

[16] Z. Su, A. Yacob, J. Wen, G. Roerink, Y. He, B. Gao, H. Boogaard, and C. van Diepen, "Assessing relative soil moisture with remote sensing data: Theory, experimental validation, and application to drought monitoring over the North China plain," *Phys. Chem. Earth, Parts A/B/C*, vol. 28, nos. 1–3, pp. 89–101, Jan. 2003.

[17] J.-F. Crétaux, A. Arsen, S. Calmant, A. Kouraev, V. Vuglinski, M. Bergé-Nguyen, M.-C. Gennero, F. Nino, R. A. D. Rio, A. Cazenave, and P. Maisongrande, "SOLS: A lake database to monitor in the near real time water level and storage variations from remote sensing data," *Adv. Space Res.*, vol. 47, no. 9, pp. 1497–1507, May 2011.

[18] Q. Feng, J. Liu, and J. Gong, "Urban flood mapping based on unmanned aerial vehicle remote sensing and random forest classifier—A case of Yuyao, China," *Water*, vol. 7, no. 12, pp. 1437–1455, Mar. 2015.

[19] E. Batur and D. Maktav, "Assessment of surface water quality by using satellite images fusion based on PCA method in the Lake Gala, Turkey," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2983–2989, May 2019.

[20] D. Lee, S. Son, H. Joo, K. Kim, M. J. Kim, H. K. Jang, M. S. Yun, C.-K. Kang, and S. H. Lee, "Estimation of the particulate organic carbon to chlorophyll—A ratio using MODIS-aqua in the East/Japan sea, South Korea," *Remote Sens.*, vol. 12, no. 5, p. 840, Mar. 2020.

[21] M. Masocha, T. Dube, T. Nhiwatiwa, and D. Choruma, "Testing utility of landsat 8 for remote assessment of water quality in two subtropical African reservoirs with contrasting trophic states," *Geocarto Int.*, vol. 33, no. 7, pp. 667–680, Jul. 2018.

[22] K. T. Peterson, V. Sagan, and J. J. Sloan, "Deep learning-based water quality estimation and anomaly detection using landsat-8/Sentinel-2 virtual constellation and cloud computing," *GISci. Remote Sens.*, vol. 57, no. 4, pp. 510–525, May 2020.

[23] B. Sharma, M. Kumar, D. M. Denis, and S. K. Singh, "Appraisal of river water quality using open-access Earth observation data set: A study of river Ganga at Allahabad (India)," *Sustain. Water Resour. Manage.*, vol. 5, no. 2, pp. 755–765, Jun. 2019.

[24] K. Ismail, A. Boudhar, A. Abdelkrim, H. Mohammed, S. Mouatassime, A. Kamal, E. Driss, E. Idrissi, and W. Nouaim, "Evaluating the potential of Sentinel-2 satellite images for water quality characterization of artificial reservoirs: The Bin El Ouidane Reservoir case study (Morocco)," *Meteorol. Hydrol. Water Manage.*, vol. 7, no. 1, pp. 31–39, Jan. 2019.

[25] C. Du, Q. Wang, Y. Li, H. Lyu, L. Zhu, Z. Zheng, S. Wen, G. Liu, and Y. Guo, "Estimation of total phosphorus concentration using a water classification method in inland water," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 71, pp. 29–42, Sep. 2018.

[26] M. Elhag, I. Gitas, A. Othman, J. Bahrawi, and P. Gikas, "Assessment of water quality parameters using temporal remote sensing spectral reflectance in arid environments, Saudi Arabia," *Water*, vol. 11, no. 3, p. 556, Mar. 2019.

[27] M. Weiss, F. Jacob, and G. Duveiller, "Remote sensing for agricultural applications: A meta-review," *Remote Sens. Environ.*, vol. 236, Jan. 2020, Art. no. 111402.

[28] E. Hancer, B. Xue, and M. Zhang, "Differential evolution for filter feature selection based on information theory and feature ranking," *Knowl.-Based Syst.*, vol. 140, pp. 103–119, Jan. 2018.

[29] Y. Zhang, D.-W. Gong, X.-Z. Gao, T. Tian, and X.-Y. Sun, "Binary differential evolution with self-learning for multi-objective feature selection," *Inf. Sci.*, vol. 507, pp. 67–85, Jan. 2020.

[30] G. Hassan, M. E. Shaheen, and S. A. Taie, "Prediction framework for water quality parameters monitoring via remote sensing," in *Proc. 1st Int. Conf. Smart Syst. Emerg. Technol. (SMARTTECH)*, Nov. 2020, pp. 59–64.

[31] Z. Lee, S. Shang, L. Qi, J. Yan, and G. Lin, "A semi-analytical scheme to estimate Secchi-disk depth from landsat-8 measurements," *Remote Sens. Environ.*, vol. 177, pp. 101–106, May 2016.

[32] P. L. Brezonik, L. G. Olmanson, J. C. Finlay, and M. E. Bauer, "Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters," *Remote Sens. Environ.*, vol. 157, pp. 199–215, Feb. 2015.

[33] M. Govedarica and G. Jakovljevic, "Monitoring spatial and temporal variation of water quality parameters using time series of open multispectral data," in *Proc. 7th Int. Conf. Remote Sens. Geoinf. Environ. (RSCy)*, Jun. 2019, Art. no. 111740.

[34] N. T. T. Ha, N. T. P. Thao, K. Koike, and M. T. Nhuan, "Selecting the best band ratio to estimate chlorophyll—A concentration in a tropical freshwater lake using sentinel 2A images from a case study of lake ba be (Northern Vietnam)," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 9, p. 290, Sep. 2017.

[35] P. R. Renosh, D. Doxaran, L. D. Keukelaere, and J. I. Gossn, "Evaluation of atmospheric correction algorithms for Sentinel-2-MSI and Sentinel-3-OLCI in highly turbid estuarine waters," *Remote Sens.*, vol. 12, no. 8, p. 1285, Apr. 2020.

[36] K. Blix, K. Pálffy, V. Tóth, and T. Eltoft, "Remote sensing of water quality parameters over Lake Balaton by using Sentinel-3 OLCI," *Water*, vol. 10, no. 10, p. 1428, Oct. 2018.

[37] W. G. Buma and S.-I. Lee, "Evaluation of Sentinel-2 and landsat 8 images for estimating chlorophyll—A concentrations in Lake Chad, Africa," *Remote Sens.*, vol. 12, no. 15, p. 2437, Jul. 2020.

[38] K. Toming, T. Kutser, A. Laas, M. Sepp, B. Paavel, and T. Nõges, "First experiences in mapping lake water quality parameters with Sentinel-2 MSI imagery," *Remote Sens.*, vol. 8, no. 8, p. 640, Aug. 2016.

[39] T. Kutser, B. Paavel, C. Verpoorter, M. Ligi, T. Soomets, K. Toming, and G. Casal, "Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters," *Remote Sens.*, vol. 8, no. 6, p. 497, Jun. 2016.

[40] Y. Zhang, R. Ma, H. Duan, S. Loiselle, M. Zhang, and J. Xu, "A novel MODIS algorithm to estimate chlorophyll a concentration in eutrophic turbid lakes," *Ecol. Indicators*, vol. 69, pp. 138–151, Oct. 2016.

[41] J. Li, Y. Zhang, R. Ma, H. Duan, S. Loiselle, K. Xue, and Q. Liang, "Satellite-based estimation of column-integrated algal biomass in nonalgae Bloom conditions: A case study of Lake Chaohu, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 2, pp. 450–462, Feb. 2017.

[42] K. Blix and T. Eltoft, "Machine learning automatic model selection algorithm for oceanic chlorophyll—A content retrieval," *Remote Sens.*, vol. 10, no. 5, p. 775, May 2018.

[43] Y. Fan, W. Li, K. J. Voss, C. K. Gatebe, and K. Stamnes, "Neural network method to correct bidirectional effects in water-leaving radiance," *Appl. Opt.*, vol. 55, no. 1, pp. 10–21, 2016.
[44] Y. Fan, W. Li, C. K. Gatebe, C. Jamet, G. Zibordi, T. Schroeder, and K. Stamnes, "Atmospheric correction over coastal waters using multilayer neural networks," *Remote Sens. Environ.*, vol. 199, pp. 218–240, Sep. 2017.
[45] E. S. El Din, Y. Zhang, and A. Suliman, "Mapping concentrations of surface water quality parameters using a novel remote sensing and artificial intelligence framework," *Int. J. Remote Sens.*, vol. 38, no. 4, pp. 1023–1042, Feb. 2017.
[46] C. Brockmann, R. Doerffer, M. Peters, S. Kerstin, S. Embacher, and A. Ruescas, "Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters," in *Proc. Living Planet Symp., Conf. Held*, vol. 740, L. Ouwehand, Ed. Prague, Czech Republic: ESA-SP, Aug. 2016, p. 54.
[47] K. Toming, T. Kutser, R. Uiboupin, A. Arikas, K. Vahter, and B. Paavel, "Mapping water quality parameters with Sentinel-3 ocean and land colour instrument imagery in the Baltic Sea," *Remote Sens.*, vol. 9, no. 10, p. 1070, Oct. 2017.
[48] A. Ansper and K. Alikas, "Retrieval of chlorophyll a from Sentinel-2 MSI data for the European union water framework directive reporting purposes," *Remote Sens.*, vol. 11, no. 1, p. 64, Dec. 2018.
[49] S. Hafeez, M. Wong, H. Ho, M. Nazeer, J. Nichol, S. Abbas, D. Tang, K. Lee, and L. Pun, "Comparison of machine learning algorithms for retrieval of water quality indicators in case-II waters: A case study of Hong Kong," *Remote Sens.*, vol. 11, no. 6, p. 617, Mar. 2019.
[50] T. Soomets, K. Uudeberg, D. Jakovels, A. Brauns, M. Zagars, and T. Kutser, "Validation and comparison of water quality products in baltic lakes using Sentinel-2 MSI and Sentinel-3 OLCI data," *Sensors*, vol. 20, no. 3, p. 742, Jan. 2020.
[51] M. Maimaitijiang, V. Sagan, P. Sidike, S. Hartling, F. Esposito, and F. B. Fritschi, "Soybean yield prediction from UAV using multimodal data fusion and deep learning," *Remote Sens. Environ.*, vol. 237, Feb. 2020, Art. no. 111599.
[52] M. E. Goher, M. H. H. Ali, and S. M. El-Sayed, "Heavy metals contents in Nasser Lake and the Nile River, Egypt: An overview," *Egyptian J. Aquatic Res.*, vol. 45, no. 4, pp. 301–312, Dec. 2019.
[53] M. E. Goher, H. I. Farhat, M. H. Abdo, and S. G. Salem, "Metal pollution assessment in the surface sediment of Lake Nasser, Egypt," *Egyptian J. Aquatic Res.*, vol. 40, no. 3, pp. 213–224, 2014.
[54] Water Environmental Federation, *Standard Methods for the Examination of Water and Wastewater*. Washington, DC, USA: American Public Health Association (APHA), 2005.
[55] *Sentinel-2 User Handbook; ESA Standard Document*, ESA Sentinel, Paris, France, 2015.
[56] A. A. Heidari and R. A. Abbaspour, "Enhanced chaotic grey wolf optimizer for real-world optimization problems: A comparative study," in *Handbook of Research on Emergent Applications of Optimization Algorithms*. IGI Global, Harrisburg, PA, USA, 2018, pp. 693–727.
[57] A. A. Heidari, H. Faris, I. Aljarah, and S. Mirjalili, "An efficient hybrid multilayer perceptron neural network with grasshopper optimization," *Soft Comput.*, vol. 23, no. 13, pp. 7941–7958, 2018.
[58] F. Liébana-Cabanillas, V. Marinković, and Z. Kalinić, "A SEM-neural network approach for predicting antecedents of M-commerce acceptance," *Int. J. Inf. Manage.*, vol. 37, no. 2, pp. 14–24, Apr. 2017.
[59] G. W.-H. Tan, K.-B. Ooi, L.-Y. Leong, and B. Lin, "Predicting the drivers of behavioral intention to use mobile learning: A hybrid SEM-neural networks approach," *Comput. Hum. Behav.*, vol. 36, pp. 198–213, Jul. 2014.
[60] J. Charvát, A. Procházka, M. Fričl, O. Vyšata, and L. Himmlová, "Diffuse reflectance spectroscopy in dental caries detection and classification," *Signal, Image Video Process.*, vol. 14, no. 5, pp. 1063–1070, 2020.
[61] J. F. Mas and J. J. Flores, "The application of artificial neural networks to the analysis of remotely sensed data," *Int. J. Remote Sens.*, vol. 29, no. 3, pp. 617–663, Feb. 2008.
[62] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Adv. Eng. Softw.*, vol. 95, pp. 51–67, May 2016.
[63] G. Hassan and A. E. Hassanien, "Retinal fundus vasculature multilevel segmentation using whale optimization algorithm," *Signal, Image Video Process.*, vol. 12, no. 2, pp. 263–270, Feb. 2018.
[64] Y. Li, J. Xu, R. Xia, X. Wang, and W. Xie, "A two-stage framework of target detection in high-resolution hyperspectral images," *Signal, Image Video Process.*, vol. 13, no. 7, pp. 1339–1346, Oct. 2019.
[65] S. M. Kloiber, P. L. Brezonik, L. G. Olmanson, and M. E. Bauer, "A procedure for regional lake water clarity assessment using landsat multispectral data," *Remote Sens. Environ.*, vol. 82, no. 1, pp. 38–47, Sep. 2002.
[66] D. Barrett and A. Frazier, "Automated method for monitoring water quality using landsat imagery," *Water*, vol. 8, no. 6, p. 257, Jun. 2016.
[67] M. A. Warren, S. G. H. Simis, V. Martinez-Vicente, K. Poser, M. Bresciani, K. Alikas, E. Spyrakos, C. Giardino, and A. Ansper, "Assessment of atmospheric correction algorithms for the Sentinel-2A multispectral imager over coastal and inland waters," *Remote Sens. Environ.*, vol. 225, pp. 267–289, May 2019.
[68] M. E. M. Hassouna, M. E. Goher, S. M. El-Sayed, and R. A. A. A. Hassan, "Integrated approach to quality indices and health risk assessment of water in the Bahr Yusuf Canal, Fayoum, Egypt," *Oceanolog. Hydrobiol. Stud.*, vol. 48, no. 4, pp. 337–354, Dec. 2019.
[69] R. Rizk, T. Juzsakova, I. Cretescu, M. Rawash, V. Sebestyén, C. Le Phuoc, Z. Kovács, E. Domokos, Á. Rédey, and H. Shafik, "Environmental assessment of physical-chemical features of Lake Nasser, Egypt," *Environ. Sci. Pollut. Res.*, vol. 27, no. 16, pp. 20136–20148, 2020.
[70] "Moroccan official bulletin. Joint orders no. 1275-01, 1276-S01 and 1277-01 of 17th October 2002 defining the quality norms of surface waters, waters destined for irrigation and of surface waters used for the production of drinking water respectively; official bulletin of the kingdom of Morocco no. 5062; Moroccan official bulletin: Rabat, Morocco," Moroccan Official Bull., Tech. Rep., 2002, pp. 1518–1525.
[71] M.-C. Tu, P. Smith, and A. M. Filippi, "Hybrid forward-selection method-based water-quality estimation via combining landsat TM, ETM+, and OLI/TIRS images and ancillary environmental data," *PLoS ONE*, vol. 13, no. 7, 2018, Art. no. e0201255.

GEHAD HASSAN received the B.Sc. degree (Hons.) in computer science from the Faculty of Computers and Information (FCI), Fayoum University, Egypt, in 2012, and the M.Sc. degree from the Faculty of Computers and Information, Cairo University, Egypt, in 2017. She is currently pursuing the Ph.D. degree with the Faculty of Computers and Information, Fayoum University. Her M.Sc. topic was about computer-aided diagnosis system for retinal images. While her Ph.D. topic is about automated water quality monitoring system based on machine learning. She is currently a Lecturer Assistant with the Faculty of Computers and Information, Fayoum University. She has published several articles in international journals and peer-reviewed international conference proceedings along with a book chapter. Her research interests include machine learning, bio-inspired optimization, image and signal processing, remote sensing, and medical image processing.

MOHAMED E. GOHER was born in September 1969. He received the Ph.D. degree from the Faculty of Science, Al Azher University, Egypt, in 2002. He is currently a Research Professor and the Head of the Chemistry Department, Freshwater and Lakes Division, National Institute of Oceanography and Fisheries (NIOF), Egypt. Since 1996, he has been working as a team leader in many research projects that have been carried out in the Nile River, Egyptian lakes, and wetland ecosystems. He is a member of many national and international councils, committees, and societies in the field of limnology, aquatic environmental sciences. He is a consultant for many international and local journals. He has contributed to producing chapters in international reference books. He is teaching postgraduate lectures on, limnology, aquatic ecology, environmental impact assessment, and lakes management. He is collaborating and consultant in the EIA studies for national organizations and projects. He is attending many training courses and scholarships as well as several national and international symposia, conferences, and congresses in Egypt and abroad. He has 50 publications in international specialized journals, covering many aspects of aquatic ecology such as water quality, risk assessment, heavy metal pollution, and wastewater treatment in Egyptian lakes, the River Nile, and wetland ecosystems.

MASOUD E. SHAHEEN received the B.Sc. degree in science from the Department of Mathematics and Computer Science, Minia University, in 1996, the M.S. degree in computer science from the Faculty of Science, Fayoum University, Egypt, in 2005, and the Ph.D. degree in computer science from the University of Southern Mississippi, in 2013. He is currently the Director of the Quality Assurance Unit, Faculty of Computers and Information, Fayoum University, where he is also a Project Portal Manager. He is also the Executive Director of Information with Fayoum University.

SHEREEN A. TAIE received the B.Sc. degree in computer science from the Faculty of Science, Cairo University, in 1996, and the M.S. and Ph.D. degrees in computer science from the Computer and Mathematics Department, Faculty of Science, Cairo University, in 2006 and 2012, respectively. She is currently the Vice Dean for Students Affairs with the Faculty of Computer and Information, Fayoum University, Egypt, where she is also an Associate Professor with the Computer Science Department. She is also the Head of the Center of Electronic Courses Production.