IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 1, February 2026, pp. 559~567
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i1.pp559-567

Evaluation of artificial intelligence algorithms to estimate water quality parameters using satellite images

Julio Cesar Anaya-Valenzuela, Gloria Yaneth Florez-Yepes, Yeison Alberto Garcés-Gómez
Faculty of Engineering and Architecture, Universidad Católica de Manizales, Manizales, Colombia

Article Info

Article history:

Received Aug 13, 2024
Revised Dec 18, 2025
Accepted Jan 10, 2026

Keywords:

Deep learning
Machine learning
Remote sensing
Spectral signature
Water quality

ABSTRACT

The Ciénaga de la Virgen (Virgen Swamp) is a coastal lagoon in Cartagena de Indias that provides multiple ecosystem services in northern Bolívar. This ecosystem has faced anthropogenic pressure from city growth and improper water resource management, including wastewater and agrochemical discharges. Consequently, environmental authorities must monitor certain sites within the water body and extrapolate the data across its entire expanse. In this study, predictive tools are applied to determine water quality parameters such as chlorophyll-a (CL-a), dissolved oxygen (DO), total suspended solids (TSS), and salinity. This is achieved by correlating traditionally obtained data with the spectral response of medium-resolution satellite images, adjusted using artificial intelligence (AI) algorithms. Support vector machine (SVM) algorithms were used for regression, random forests (RF), and artificial neural networks (ANN), achieving an accuracy of 79% for CL-a, 95% for DO, 89% for TSS, and 96% for salinity. Validation was performed using mean absolute percentage error (MAPE) statistical metrics and root mean square error (RMSE).

Corresponding Author:

Yeison Alberto Garcés-Gómez
Master in Remote Sensing, Faculty of Engineering and Architecture, Universidad Católica de Manizales
Cra 23 No 60-63, Manizales, Caldas, Colombia
Email: ygarces@ucm.edu.co

1. INTRODUCTION

Remote sensing has proven to be an important technological tool in capturing information from the earth's surface [1], aiding in the management and sustainable use of natural resources. Optical satellite images contribute to the monitoring of water resources [2], providing a support instrument that, along with timely captured data, allows the modeling of the behavior of water quality variables. In this article, artificial intelligence (AI) algorithms are applied and adjusted with field measurements [3] to calculate the water quality parameters of chlorophyll-a (CL-a), dissolved oxygen (DO), total suspended solids (TSS), and salinity. The study utilizes satellite images of medium spatial resolution from the Ciénaga de la Virgen (Virgen Swamp), located in the city of Cartagena de Indias in northern Bolívar. This body of water sustains many families living in its area of influence [4], who are affected by the deterioration of the waters due to anthropic pressure [5], mainly caused by inadequate sewage discharges and solid waste, among other problems of this lentic body. A comparison will be made in the performance of the proposed AI algorithms, including support vector machines (SVM) for regression, random forests (RF) (classified under machine learning (ML) methods), and artificial neural networks (ANN) (part of deep learning (DL)) [6]. This study focuses on calculating the continuous variables of CL-a, TSS, and salinity, chosen for their optical properties [7]. DO is also included, selected for its critical role as an indicator of the aquatic ecosystem's capacity to support flora and fauna [6]. The estimation of these parameters through algorithms serves to complement the

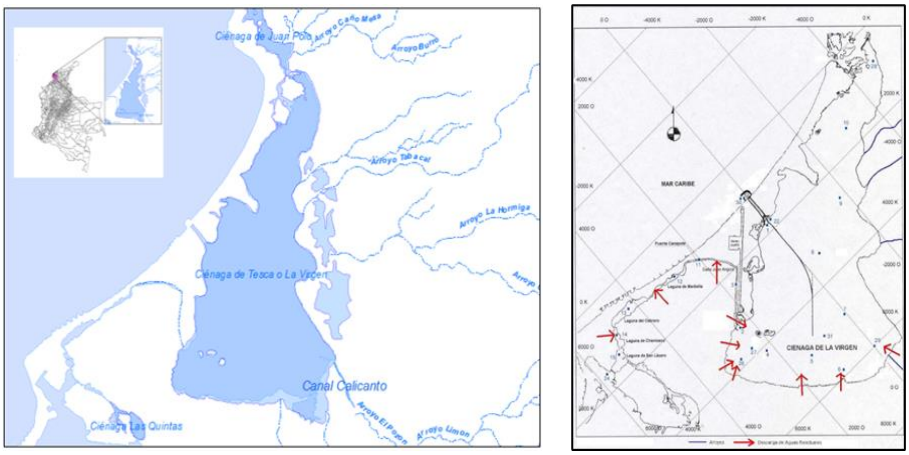
management of water resources, without intending to replace traditional methodologies that are globally standardized and certified [8], conducted by specialized laboratories in water quality measurement.

Furthermore, recent advancements in ML offer new avenues for enhancing model robustness, particularly when dealing with limited or noisy datasets. Methodologies such as self-supervised learning (SSL) show promise in improving performance by leveraging the large amounts of unlabeled satellite data available. Additionally, the integration of data through multi-sensor fusion techniques may provide a more comprehensive spectral understanding, potentially improving the accuracy of water quality predictions over a reliance on a single data source. This study aims to evaluate the performance of various AI algorithms. The evaluation focuses on estimating CL-a, DO, TSS, and salinity parameters as indicators of water quality in the Virgen Swamp. This evaluation will utilize medium spatial resolution satellite images obtained from the Google Earth Engine and Jupyter Notebook platform.

2. METHOD

2.1. Establishment of the area of interest and sampling points

The Virgen Swamp is situated in the city of Cartagena de Indias, in the Department of Bolívar, covering a total area of 502.45 km². It is a coastal lagoon separated from the sea by a sand barrier between 400 and 800 meters wide that starts in the village of La Boquilla. It has a triangular shape, with a width to the south of 4.5 km and a length of 7 km. The location and general characteristics of the study area are presented in Figure 1. In particular, the body of water of the Ciénaga de la Virgen measures approximately 22.5 km^2 and has depths of up to 1.6 m [9] as shown in Figure 1(a). In order to make use of the information provided, it was necessary to georeference the location plan provided, since the data did not have the exact coordinate of the sampling in the field. The georeferencing of the plane was conducted to closely match the geometry of the body of water using identifiable sinuosity on the shoreline. The plane's grid uses arbitrary coordinates, which posed challenges in achieving precise georeferencing. This limitation affected the positional accuracy of the sampling points, leading to uncertainties in the prediction models. After georeferencing the plan, the laboratory digitized 10 sampling sites within the body of water identified by numbers: 2, 4, 5, 6, 7, 8, 10, 22, 28, and 32. Figure 1(b) shows the location plan provided by the CARDIQUE laboratory. To ensure the experimental setup is clear for replication, the georeferencing process, despite its challenges, was conducted as follows. The scanned laboratory plan, as shown in Figure 1(b), was imported into a geographic information system (GIS) environment. Identifiable shoreline features and sinuosity visible in both the plan and baseline satellite imagery were used as ground control points to align the plan's arbitrary grid to the real-world coordinate system. Although this manual alignment introduces positional uncertainty, the 10 digitized sampling points represent the best available approximation of the in-situ collection sites.



(a) (b)

Figure 1. Location of the study area of (a) Virgen Swamp and (b) sampling points in the swamp

2.2. Verification and adjustment of existing information of the area of interest

The regional autonomous corporation of the Canal del Dique-Cardique provided the results of water quality analysis for the Virgen Swamp from 2015 to 2021. The data comprise a total of 37 sampling records from both the body of water and the channels that feed into the swamp. The information has an identifier per point that is listed on a drawing provided by the CARDIQUE laboratory.

2.3. Selection of satellite images

The dates of the filtered images match the dates of field sample collection by the laboratory. This approach ensured that the spectral response of the sites of interest corresponded closely to the water conditions analyzed by the laboratory [10], thereby minimizing uncertainty related to the dynamics of the swamp's water body. It is crucial to consider the time gap between image capture and laboratory sampling, as water conditions are subject to constant changes influenced by factors such as discharges, temperature variations, unexpected rainfall, and other anthropogenic or environmental conditions in the area [11].

Out of the 37 field sampling dates, 7 images or image collections were found to have coinciding or closely matching capture dates. These images or image collections have a spatial resolution of 3 meters, a radiometric resolution of 16 bits, are orthorectified, and corrected to surface reflectance [12], facilitating direct analysis. A total of seven image capture dates were identified that closely aligned with the laboratory sampling dates, ensuring temporal consistency between the remote sensing data and the field measurements. This correspondence is critical for minimizing uncertainty related to the dynamic nature of the swamp's water body. The specific pairings of satellite image dates and laboratory collection dates utilized for this study are detailed in Table 1.

Table 1. Date of capture of the satellite image versus date of collection of the sample by the laboratory

Satellite image date	Date of sample collection by the laboratory
06/20/2017	06/21/2017
07/22/2017	07/27/2017
09/2/2017	08/31/2017
09/11/2017	09/28/2017
07/22/2018	07/23/2018
08/21/2018	08/21/2018
03/26/2019	03/27/2019

2.4. Preparing and uploading information to the Google Earth Engine platform

The Google Earth Engine is a cloud-based platform designed for geospatial data analysis, widely utilized by researchers globally for trend analysis [13]. It leverages Google's robust computing infrastructure to support various research domains [14]. To determine the area of interest, the initial step involved using the geometry in shapefile format, specifically the detailed permanent channel of the Virgen Swamp. This data was derived from technical studies conducted by CARDIQUE in 2021, which delineated the water perimeter of the swamp and internal water bodies within Cartagena. The geometry was uploaded to the Google Earth Engine platform as assets along with the point geometry with the laboratory information in shapefile format and the selected image collections with take date 06/20/2017, 07/22/2017, 09/2/2017, 09/11/2017, 07/22/2018, 08/21/2018, and 03/26/2019.

2.5. Obtaining numerical models

To determine the algorithms correlating CL-a, DO, TSS, and salinity concentrations, the study selected four image bands: b1 (blue, 0.455-0.515 µm), b2 (green, 0.5-0.59 µm), b3 (red, 0.59-0.67 µm), and b4 (near-infrared (NIR), 0.78-0.86 µm). These bands are commonly used in water body studies [2] and were chosen as independent or regression variables. According to Briceño *et al.* [15], as wavelength increases, water absorbs more incident energy, resulting in lower or negligible energy reflection beyond the NIR bands, which therefore do not contribute significantly to water quality analysis [16]. Likewise, the study considered the normalized difference vegetation index (NDVI), the normalized difference water index (NDWI), $\frac{b1(0.455-0.515\ \mu m)}{b2(0.5-0.59\ \mu m)}$, $\frac{b2(0.5-0.59\ \mu m)}{b3(0.59-0.67\ \mu m)}$, $\frac{b2(0.5-0.59\ \mu m)}{b4(0.78-0.86\ \mu m)}$, $\frac{b3(0.59-0.67\ \mu m)}{b4(0.78-0.86\ \mu m)}$ and simple ratios based on spectral signature analysis at different levels of incident energy, as conducted by Ruddick *et al.* [17]. For the calculation of CL-a, prominent peaks in the green region (b2, 0.5-0.59 µm) and energy absorption in the blue (b1, 0.455-0.515 µm) and red (b3, 0.59-0.67 µm) regions are observed in the spectral signature. For TSS, absorption is noticeable in the blue band (b1, 0.455-0.515 µm), increases in the green band (b2, 0.5-0.59 µm), peaks in the red band (b3, 0.59-0.67 µm), and decreases in the NIR band (b4, 0.78-0.86 µm) [18]. Additionally, new bands were selected from the principal components to leverage their low correlation. With the geometry corresponding to the field sampling sites, previously loaded as an asset in Google Earth Engine, reflectance values were extracted for each of the independent variables. The following table lists a sample of the data obtained.

2.6. Data analysis

For data analysis, AI techniques were employed using ML models with regression, including SVM, RF, and DL models such as ANN. The selection of bands, indices, and simple quotients was based on literature sources such as the study by Briceño *et al.* [15]. Their analysis of absorption and reflection levels in spectral

signatures defined these variables as explanatory factors. Similarly, based on the research results of [19], with multiple regression models whose predictor variables are based on the visible spectrum bands of the CBERS-2B satellite. Regarding the bands derived from principal components analysis (PCA), the objective is to determine whether incorporating the variability contributed by these bands optimizes models, as demonstrated in previous studies like that of Lopes *et al.* [20]. It's crucial to note that chlorophyll absorbs more electromagnetic energy in the blue band, with an increase in the green band [21]. Therefore, these bands are essential for inclusion in numerical models. Additionally, the red and infrared bands are significant for detecting suspended solids, as discussed by [22].

3. RESULTS AND DISCUSSION

3.1. Result of the model with support vector machine

The first AI algorithm utilized was SVM for regression. To perform the regression by means of this algorithm, it is necessary to have a training data set and a test data set [3], so in this study 30% of the total data was established for the test data set. This percentage is used for model prediction and validation. The scikit-learn Python library provides tools that simplify programming and mathematical calculations for models, including the SVM library, which includes the support vector regression (SVR) algorithm. Then, it is necessary to define a kernel function, which can be linear or non-linear. For this particular case, the default kernel found, radial basis function (RBF), which corresponds to a Gaussian kernel, was selected. These parameters were applied to the four water quality variables of interest. It is important to mention that the data were previously standardized, since AI algorithms can have erroneous performance if the data do not follow a more or less normal or Gaussian distribution, for this the scikit-learn tool was used, StandardScaler which eliminates the mean of the data and scales them with variance equal to 1, by means of the calculation $z = (x - u)/s$, where x it is the value of the training data, u it is the mean of the training sample, s the standard deviation [23]. After running the models, it is important to revert to the initial values to obtain data according to the original scale of the variables being predicted. This step is accomplished using the `inverse_transform` function, which is also incorporated into the StandardScaler algorithm. To calculate the accuracy of the model with SVR, the mean absolute percentage error (MAPE) was used Aguilar and Díaz [3]. MAPE is derived from the mean of the absolute percentage error (APE), which is calculated using (1).

$$APE = 100 * \left(\frac{abs(y - y')}{y}\right) \quad (1)$$

Where y is the observed value and y' prediction value. With the MAPE, the prediction of the model was calculated by means of (2).

$$Precisión = 100 - MAPE \quad (2)$$

The model with SVR allowed obtaining the following performance for each water quality parameter of interest (see Table 2). It is important to mention that the basis for the selection of predictor variables was the significant correlation between these and the response variable. However, from this basis, we proceeded with the entry of other variables empirically or the exclusion of them, seeking those that improved the performance of the model. While the initial selection of predictor variables was based on significant correlations identified in the literature and preliminary analysis, the final variable sets were refined empirically. This iterative approach was necessary to optimize model performance for this specific dataset. However, we acknowledge this empirical refinement carries a risk of overfitting and may influence the model's generalizability. A more theoretically grounded feature selection process is recommended for future studies with larger datasets.

Table 2. Model accuracy by water quality parameter with SVM for regression

Water quality parameter	Predictor variables	Accuracy (%)	MAPE (%)	RMSE
CL-a	'r_b1', 'r_b2', 'r_b3', 'r_b2_b3', 'r_ndwi'	68	32	7.2 µg/L
DO	'r_b4','r_ndwi', 'r_pc1','r_pc2','r_pc4'	93	7	0.6 mg O2/L
TSS	'r_b1','r_b3', 'r_b1_b2', 'r_b3_b4', 'r_b2_b4', 'r_ndvi'	84	16	18.5 mg/L
Salinity	'r_b1','r_b2','r_b4', 'r_b1_b2', 'r_b2_b3', 'r_ndvi', 'r_pc1'	94	6	1.9 o/oo

In order to analyze the performance of the models with SVM for regression, dispersion diagrams were generated between the observed values and the predicted values. These diagrams report a grouping of the data with a linearity pattern, especially in those whose performance exceeded 90%, as is the case with DO

and salinity (see Figure 2). According to Figure 2, and as evidenced by the analytical result of the prediction, the CL-a parameter presented the lowest performance when predicting with data different from those used in the training set. However, the accuracy of 68% agrees with the result of Ledesma *et al.* [24], which takes into account green and near-infrared as regressive variables.

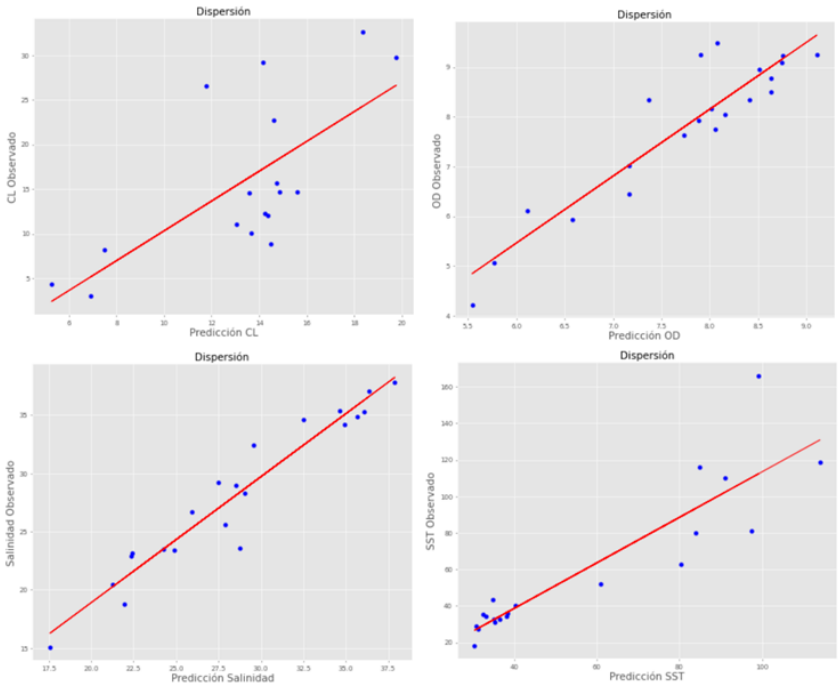


Figure 2. Dispersion diagrams between observation and prediction with SVM

3.2. Result of the model with random forests

The second model used is the RF. As in the SVM, 30% of the total sample was reserved for the validation of the model. From the Python library scikit-learn, the RF regressor algorithm is used, which has as a parameter of special importance the number of estimators or trees that make up the forest to be built. In addition to the aforementioned parameter, the RF regressor library includes a series of other parameters, such as the criterion for separability or the depth of the trees [23], among others, which were not used for the prediction of the water quality indicators proposed in this document. Before generating the forest, we proceeded with the standardization of the data, as was done with the SVRs. After the prediction, it was necessary to rescale the data to the initial values of the water quality parameters of interest in order to obtain comparable information and verify the accuracy of the model. This accuracy was calculated in the same way using the MAPE. Modeling with RF presents a significant improvement with respect to SVM, with a marked linear trend between observation and prediction as shown in Table 3.

Table 3. Model accuracy by water quality parameter with RF for regression

Water quality parameter	Predictor variables	Accuracy (%)	MAPE (%)	RMSE
CL-a	'r_b3', 'r_b1_b2'	79	21	2.7 µg/L
DO	'r_b4', 'r_b2_b4','r_pc4'	95	5	0.5 mg O2/L
TSS	'r_b1','r_b2','r_ndwi'	89	11	8.9 mg/L
Salinity	'r_b1','r_b2_b3','r_b2_b4'	96	4	1.1 o/oo

3.3. Result of the model with artificial neural networks

The last model applied corresponds to the ANN and in the same way as in the two previous models, 30% of the sample data was reserved, in order to use them as test or test data for model validation. To work with DL in Python, it is essential to utilize libraries such as Theano, TensorFlow, and Keras. These libraries are designed for optimizing numerical models and provide tools that simplify the development of neural networks. In the code, when Keras is imported by default, Tensorflow is loaded, and within Keras there are two modules necessary for the construction of the red neuronal artificial (RNA) model. The first module is sequential, which allows you to initialize the network parameters, and the second module is dense, which

allows you to create the intermediate layers of the network. In the dense module, the units parameter was utilized, representing the number of nodes in the hidden layer. This is an empirical hyperparameter determined by the developer. Additionally, the input_dim parameter, which specifies the number of independent variables, was replaced with the expression X_train.shape [1], a command that returns the total number of explanatory variables. In the kernel_initializer, the function that initializes the weights of the feature array is entered randomly, for which a normal distribution function was used and the activation function allows determining by means of the weights whether or not the information follows the next layer according to its importance.

After configuring the first layer, the second hidden layer was added with parameters similar to those of the previous layer, except for input_dim, which was already specified in the first layer. In this case, the activation function was changed to tanh. Subsequently, the output layer was defined with a single unit and a normal distribution initialization kernel. To compile the model, a loss function is required. The least squares error function was selected to minimize the difference between the observations and predictions. For optimization, the Adam algorithm was used. This optimizer helps in finding the most accurate set of weights and is the default optimizer within the compilation module. Finally, the model must be trained. For this, you need the training data for the independent variables and the dependent variable vector. Additionally, you must specify the number of epochs, which denotes how many times the model will iterate over the entire dataset, and the batch size, which indicates the number of training samples processed before the model's internal parameters are updated. With the ANN, the performance summarized in Table 4 was obtained.

Table 4. Model accuracy by water quality parameter with ANN

Water quality parameter	Predictor variables	Accuracy (%)	MAPE (%)	RMSE
CL-a	'r_b1','r_b3', 'r_b1_b2', 'r_b3_b4', 'r_b2_b4', 'r_ndvi'	50	50	10.2 µg/L
DO	'r_b4', 'r_b3_b4','r_pc1','r_pc4'	87	13	1.2 mg O2/L
TSS	'r_b1','r_b2','r_ndwi'	73	27	23.6 mg/L
Salinity	'r_b1', 'r_b3', 'r_b2_b3','r_b3_b4', 'r_b2_b4', 'r_ndvi', 'r_ndwi', 'r_pc1'	90	10	3.6 o/oo

CL-a was the water quality parameter with the lowest performance using the ANN algorithm, with an estimated accuracy of 50%. This implies that the model accounts for only 50% of the total variability in the data. Although high performance was not achieved for CL-a, RF yielded a good performance with an accuracy of 79%. This falls within the precision range reported in studies of this variable using remote sensing data, which indicate an R^2 between 0.5 and 0.9 [15]. Additionally, Aguilar and Díaz [3] reported that AI techniques can achieve accuracies greater than 60%. This result was achieved using the red band and the simple quotient between the blue and green bands. This aligns with several studies [15], [19], [21], among others, that associate the visible spectrum with the water quality parameter of CL-a. Phytoplankton exhibits an absorption peak in the blue band [25], with wavelengths between 0.455 and 0.515 µm, and a reflection maximum in the green band, with wavelengths between 0.5 and 0.59 µm, followed by a decline in wavelengths greater than 0.59 µm.

DO, similar to CL-a, demonstrated its highest performance with the RF algorithm, achieving an estimated accuracy of 95%. The predictor variables for this model include the near infrared band, the simple quotient between the green and near infrared bands, and main component 4. The green band is particularly noteworthy, as it shows significant results in regression models, such as the one by Alzate *et al*. [26] with an R^2 of 0.77, and the simple linear regression model from Vanegas [27] with an R^2 of 0.8. Therefore, this result could be valuable for predicting DO levels in the Virgen Swamp. Vanegas [27], in his results specified that he did not find a relationship between the analyzed parameters and the near-infrared, which differs from what is presented here since this band contributed in the prediction, not only in the DO where a Spearman correlation coefficient of 0.3 was obtained with a p-value of 0.018 lower than the statistical significance α=0.05, which allowed rejecting the null hypothesis that considers that the near-infrared band does not correlate with the DO parameter, but also contributed in the prediction of the other parameters studied.

TSS demonstrated strong performance across all three proposed models. The most significant results were achieved with the RF model, which had an accuracy of 89%. The key predictor variables included the blue and green bands, as well as the NDWI, which is derived from the normalized difference between the green and NIR bands. These bands were also crucial for predicting TSS in the study by [28], where the best predictions were obtained using the quotient of these variables. This model also showed the smallest mean squared error (MSE) and MAPE, consistent with the metrics used in this paper. These bands have been employed in various studies to determine TSS, often in conjunction with simple quotients. For instance, Gholizadeh *et al*. [29] utilized these bands with models incorporating DL techniques, while [30] applied simple regressions using the visible spectrum bands from the TM sensor on the Landsat 5 platform.

However, the latter reference reported a low performance for their model. RF have contributed with satisfactory results in other studies for occupational safety and health (OSH) modelling such as the one presented by [31], although the correlation is made with other physicochemical parameters of water quality.

The best model for representing salinity was the RF algorithm, achieving an accuracy of 96%. This model outperformed the SVM models, which had an accuracy of 94%, and the ANN, which achieved an accuracy of 90%. The latter showed favorable results compared to multiple linear regression models, as noted by Zhou *et al.* [32], and demonstrated accuracies equal to or exceeding 90%, as reported in the salinity variation study by Huang and Foo [33]. Salinity was the parameter with the best performance throughout the analysis. This was evidenced by a Spearman correlation coefficient of 0.6 and a p-value of less than 0.05, which were obtained from the direct analysis of the simple quotient between the blue band (0.455 to 0.515 µm) and the green band (0.5 to 0.59 µm).

Numerical modeling with information acquired through remote sensing has demonstrated its usefulness in predicting water quality parameters [21] such as those selected in this study. The performance of the models is influenced by factors such as the time difference between when the laboratory sample is collected and when the remote sensor image is captured. Therefore, it is crucial to ensure that the data are collected at the same time. As noted by Bazán *et al.* [21], if there are temporal discrepancies, it is essential to verify that the water body has not experienced disturbances, such as precipitation or contaminant discharges, that could alter the spectral response of the satellite image or introduce anomalous values or outliers in the laboratory sample. These outliers should be excluded from the sample to avoid negatively impacting the algorithm's performance. It is important to acknowledge the limitations of this study, primarily the small sample size. The dataset consisted of only 37 sampling records, which constrained the model training and validation process, especially for complex algorithms like ANN. Consequently, the validation was performed using a simple 70/30 train-test split. While this provided initial performance metrics, the robustness of the models could be further improved by employing more rigorous techniques, such as k-fold cross-validation, which is better suited for limited datasets. Future work should aim to incorporate a larger dataset to validate and enhance the generalizability of these findings.

For future research, exploring more advanced DL architectures could yield significant improvements. For instance, transformer-based deep networks, which have shown success in noise reduction for other domains like medical imaging, could be adapted for processing satellite imagery. Such models might prove effective in mitigating atmospheric interference and other noise inherent in remote sensing data, thereby improving the quality of spectral signatures used for water quality estimation.

4. CONCLUSION

The use of satellite imagery combined with AI models constitutes an effective complementary tool for monitoring water quality in lentic bodies such as the Virgen Swamp, as it reveals significant relationships between spectral reflectance and parameters such as CL-a, DO, salinity, and TSS, although it does not replace laboratory analyses. ML and DL models showed advantages over traditional statistical approaches by not relying on parametric assumptions and by achieving high levels of accuracy, with salinity standing out as the best-performing parameter (up to 96% with RF), while CL-a presented the lowest performance, possibly due to the influence of the swamp bed. Overall, the results confirm that these models enable efficient, low-cost monitoring with potential as an early warning system for environmental management, supported by open platforms such as Google Earth Engine and Jupyter Notebook, and provide valuable information for decision-making by environmental authorities.

FUNDING INFORMATION

Authors state no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Julio Cesar Anaya-Valenzuela	✓	✓	✓		✓	✓	✓	✓	✓				✓	✓
Gloria Yaneth Florez-Yepes	✓	✓		✓		✓	✓	✓	✓			✓	✓	✓
Yeison Alberto Garcés-Gómez		✓		✓		✓	✓			✓	✓	✓		✓

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, [YAGG], upon reasonable request.

REFERENCES

[1] E. Chuvieco, *Fundamentals of satellite remote sensing an environmental approach*. New York, United States: CRC Press, Taylor & Francis Group, 2020.

[2] J. A.- Alvarez, P. P.- Cutillas, L. C. A.- Cejudo, and O. R.- Valle, "Multispectral analysis for estimating turbidity as an indicator of water quality in reservoirs in the state of Chihuahua, Mexico (in Spanish: *Análisis multiespectral para la estimación de la turbidez como indicador de la calidad del agua en embalses del estado de Chihuahua, México*)," *Revista Geográfica de América Central*, vol. 1, no. 62, pp. 33–61, Sep. 2018, doi: 10.15359/rgac.62-1.2.

[3] A. C. A. Aguilar and F. F. O.- Díaz, "Machine learning for predicting drinking water quality," *Ingeniare*, no. 28, pp. 47–62, 2020, doi: 10.18041/1909-2458/ingeniare.28.6215.

[4] W. Maldonado, I. Baldiris, and J. Díaz, "Assessment of water quality of Ciénaga de la Virgen (Cartagena, Colombia) during the period 2006-2010," *Revista Científica Guillermo de Ockham*, vol. 9, no. 2, pp. 79–87, Dec. 2011.

[5] Cardique, *Plan for the management and conservation of the Cienaga de la Virgen watershed (in Spanish: Plan de Ordenamiento y Manejo de la Cuenca Hidrografica de la Cienaga de la Virgen)*. Bogotá, Colombia: Cardique-Conservacion International Colombia, 2019. [Online]. Available: https://es.scribd.com/doc/7096498/01-Plan-to-y-Manejo-Cuenca-Cienaga-de-La-Virgen

[6] R. Huang, C. Ma, J. Ma, X. Huangfu, and Q. He, "Machine learning in natural and engineered water systems," *Water Research*, vol. 205, Oct. 2021, doi: 10.1016/j.watres.2021.117666.

[7] A. Gitelson, G. Garbuzov, F. Szilagyi, K. H. Mittenzwey, A. Karnieli, and A. Kaiser, "Quantitative remote sensing methods for real-time monitoring of inland waters quality," *International Journal of Remote Sensing*, vol. 14, no. 7, pp. 1269–1295, 1993, doi: 10.1080/01431169308953956.

[8] R. Baird and L. Bridgewater, *Standard methods for the examination of water and wastewater standard methods for the examination of water and wastewater*, 23rd ed., no. 1. Washington, D. C, United States: American Public Health Association, 2017.

[9] Observatorio Ambiental De Cartagena de Indias, "Ciénaga de La Virgen," *observatorio.epacartagena.gov.co*. Accessed: Jan. 20, 2025. [Online]. Available: https://observatorio.epacartagena.gov.co/gestion-ambiental/ecosistemas/proyecto-cienaga-de-la-virgen/cienaga-de-la-virgen/

[10] Y. Oyama, B. Matsushita, T. Fukushima, K. Matsushige, and A. Imai, "Application of spectral decomposition algorithm for mapping water quality in a turbid lake (Lake Kasumigaura, Japan) from Landsat TM data," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 73–85, Jan. 2009, doi: 10.1016/j.isprsjprs.2008.04.005.

[11] R. M. McCoy, *Field methods in remote sensing*. New York, United States: The Guilford Press, 2005.

[12] Planet, "Planet products: real-time satellite monitoring with planet," *planet.com*. Accessed: Jan. 20, 2025. [Online]. Available: https://www.planet.com/products/satellite-monitoring/

[13] O. O. Diaz, *Basic introduction to Google Earth Engine (GEE) (in Spanish: Introducción básica a Google Earth engine (GEE))*. Bonn, Germany: Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), 2018.

[14] J. Xiong *et al.*, "Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on Google Earth Engine," *Remote Sensing*, vol. 9, no. 10, Oct. 2017, doi: 10.3390/rs9101065.

[15] I. Briceño, W. Pérez, D. S. Miguel, and S. Ramos, "Determination of water quality Vichuquén Lake, using satellite images Landsat 8, sensor OLI, year 2016, Chile," *Revista de Teledetección*, no. 52, pp. 67–78, 2018, doi: 10.4995/raet.2018.10126.

[16] C. Pohl *et al.*, *Principles of remote sensing an introductory textbook*. Enschede, Netherland: The International Institute for Aerospace Survey and Earth Sciences (ITC), 2001.

[17] K. G. Ruddick, V. De Cauwer, Y. J. Park, and G. Moore, "Seaborne measurements of near infrared water-leaving reflectance: the similarity spectrum for turbid waters," *Limnology and Oceanography*, vol. 51, no. 2, pp. 1167–1179, 2006, doi: 10.4319/lo.2006.51.2.1167.

[18] J. T.- Pérez and A. McCullum, "Remote sensing of coastal ecosystems," NASA Applied Remote Sensing Training Program (ARSET). Accessed: Jan. 20, 2020. [Online.] Available: https://appliedsciences.nasa.gov/remote-sensing-coastal-ecosystems

[19] M. Bonansea, C. Ledesma, C. Rodriguez, and A. R. S. Delgado, "Chlorophyll-a concentration and photic zone boundary in the Río Tercero reservoir (Argentina) using CBERS-2B satellite images (in Spanish: *Concentración de clorofila-a y límite de zona fótica en el embalse Río Tercero (Argentina) utilizando imágenes del satélite CBERS-2B*)," *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, vol. 7, no. 3, pp. 61–71, Dec. 2012, doi: 10.4136/ambi-agua.847.

[20] J. W. B. Lopes, F. B. Lopes, E. M. de Andrade, L. C. G. Chaves, and M. G. R. Carneiro, "Spectral response of water under different concentrations of suspended sediment: measurement and simplified modeling," *Journal of Agricultural Science*, vol. 11, no. 3, Feb. 2019, doi: 10.5539/jas.v11n3p327.

[21] R. Bazán *et al.*, "Remote sensing and numerical modeling for water quality analysis of the Los Molinos reservoir," *Ingeniería hidráulica en México*, vol. 20, no. 2, pp. 121–136, 2005.

[22] J. W. M. C. Santos and V. Dubreuil, "Estimation of the temporal and spatial distribution of suspended material in the waters of the Manso-MT reservoir based on Landsat images and field data (in Portuguese: *Estimativa da distribuição temporo-espacial de material em suspensão nas águas do reservatório de manso-mt a partir de imagens landsat e dados de campo*)," *Anais XIV Simposio Brasileiro de Sensoriamento Remoto, Natal, Brasil*, pp. 5421–5428, 2009.
[23] F. Pedregosa, G. Varoquaux, A. Gramfort, and V. Michel, "History," *scikit-learn.org*. Accessed: Jan. 20, 2025. [Online]. Available: https://scikit-learn.org/stable/about.html
[24] C. Ledesma, M. Bonansea, C. Rodríguez, and Á. R. S. Delgado, "Water quality control in third river reservoir (Argentina) using geographical information systems and linear regression models," *Ambiente e Agua - An Interdisciplinary Journal of Applied Science*, vol. 8, no. 2, pp. 67–76, 2013, doi: 10.4136/ambi-agua.1113.
[25] J. P. Cannizzaro and K. L. Carder, "Estimating chlorophyll a concentrations from remote-sensing reflectance in optically shallow waters," *Remote Sensing of Environment*, vol. 101, no. 1, pp. 13–24, Mar. 2006, doi: 10.1016/j.rse.2005.12.002.
[26] D. F. C.- Alzate, Y. A. G.- Gomez, and V. H.- Cespedes, "Landsat-7 ETM+ based remote sensing as a tool for assessing lakes water quality characteristics," *Journal of Southwest Jiaotong University*, vol. 56, no. 1, 2021, doi: 10.35741/issn.0258-2724.56.1.28.
[27] A. P. Vanegas, "Prediction of physical and chemical parameters of water quality using remote sensors: case study of the Neusa reservoir," *M.Sc. Thesis*, Maestría en Ciencias Ambientales, Facultad de Ciencias Naturales e Ingeniería, Jorge Tadeo Lozano University Foundation of Bogota, Colombia, 2025.
[28] D. C. R. Ramírez, "Method for estimating total suspended solids as an indicator of water quality using satellite imagery (in Spanish: *Método de estimación de sólidos suspendidos totales como indicador de la calidad del agua mediante imágenes satelitales*)," *M.Sc. Thesis*, Facultad de Ciencias Agrarias, Escuela de posgrado, Universidad Nacional de Colombia, Bogotá, Colombia, 2017.
[29] M. Gholizadeh, A. Melesse, and L. Reddi, "A comprehensive review on water quality parameters estimation using remote sensing techniques," *Sensors*, vol. 16, no. 8, Aug. 2016, doi: 10.3390/s16081298.
[30] A. Kulkarni, "Water quality retrieval from Landsat TM imagery," *Procedia Computer Science*, vol. 6, pp. 475–480, 2011, doi: 10.1016/j.procs.2011.08.088.
[31] A. S. Qambar and M. M. M. Al Khalidy, "Development of local and global wastewater biochemical oxygen demand real-time prediction models using supervised machine learning algorithms," *Engineering Applications of Artificial Intelligence*, vol. 118, Feb. 2023, doi: 10.1016/j.engappai.2022.105709.
[32] F. Zhou, B. Liu, and K. Duan, "Coupling wavelet transform and artificial neural network for forecasting estuarine salinity," *Journal of Hydrology*, vol. 588, 2020, doi: 10.1016/j.jhydrol.2020.125127.
[33] W. Huang and S. Foo, "Neural network modeling of salinity variation in Apalachicola River," *Water Research*, vol. 36, no. 1, pp. 356–362, Jan. 2002, doi: 10.1016/S0043-1354(01)00195-6.

BIOGRAPHIES OF AUTHORS

Julio Cesar Anaya-Valenzuela is a cadastral Engineer and Geodesist from the Universidad Distrital Francisco José de Caldas, specialist in engineering project management from the same university and master in remote sensing from the Universidad Católica de Manizales, Colombia. He works as a public servant in the governor's office of Atlántico, Colombia, on issues associated with risk management and climate change. His area of interest includes remote sensing applied to risk management and environment, geographic information systems and artificial intelligence algorithms. He can be contacted at email: julio.anaya@ucm.edu.co.

Gloria Yaneth Florez-Yepes development and the environment, Ph.D. Sustainable Development, Universidad de Manizales, Manizales, in 2018. She is professor at the Universidad Católica de Manizales Colombia, coordinator in the research group on Technological and Environmental Development. She published more than 25 scientific and research publications. She can be contacted at email: gyflorez@ucm.edu.co.

Yeison Alberto Garcés-Gómez received bachelor's degree in Electronic Engineering, and master's degrees and Ph.D. in Engineering from Electrical, Electronic and Computer Engineering Department, Universidad Nacional de Colombia, Manizales, Colombia, in 2009, 2011 and 2015, respectively. He is full professor at the Academic Unit for Training in Natural Sciences and Mathematics, Universidad Católica de Manizales, and teaches several courses such as experimental design, statistics and physics. His main research focus is on applied technologies, embedded system, power electronics, power quality, but also many other areas of electronics, signal processing and didactics. He published more than 30 scientific and research publications, among them more than 10 journal papers. He worked as principal researcher on commercial projects and projects by the Ministry of Science, Tech and Innovation, Republic of Colombia. He can be contacted at email: ygarces@ucm.edu.co.