

DOI: 10.36108/esj/5202.40.0150

Predictive Machine Learning Modelling of Biological Indices from Physico-Chemical Characteristics in Sand Dredging Areas of Lagos Lagoon

Manuscript Received: 15/09/2025 Manuscript Accepted: 18/12/2025 Article Published: 30/12/2025

AJAYI Aderemi Pius¹, SHOGA Aramide Mutiat¹, OSHONEYE Jacob oluwaseyi¹, ADEYEMO Lateef Taiwo¹, OLADAPO Kolade Paul-Preacher²

¹Environmental toxicology and pollution management unit, Zoology Department, University of Lagos, Akoka, Yaba, Lagos, Nigeria

²Applied Entomology unit, Department of zoology, Faculty of Life Science, University of Lagos

*Corresponding Author- Email: remipius@yahoo.com Orcid No: <https://orcid.org/0009-0006-3090-9292>



ABSTRACT

Sand dredging in the Lagos Lagoon Complex poses significant threats to aquatic ecosystem integrity. This study applied three machine learning (ML) regression algorithms; Random Forest (RF), K-Nearest Neighbour (KNN), and Adaptive Boosting (AdaBoost) to model and predict biological indices from physicochemical water quality variables across four sampling locations (Ibese, Majidun, Makoko, and Ilaje) in active and reference sand dredging zones of the Lagos Lagoon. Physicochemical parameters, antioxidant enzyme activities (CAT, SOD, GST, GPx) and MDA, a lipid peroxidation marker in *Chrysichthys nigrodigitatus* and *Ethmalosa fimbriata*, and biological diversity indices (Shannon–Wiener H', BMWP) derived from phytoplankton and benthic macrofauna were assessed. The VIF screening reduced 13 physicochemical predictors to four non-collinear variables: dissolved oxygen (DO), total dissolved solids (TDS), nitrate, and phosphate. The RF achieved the strongest prediction of benthic diversity (H' Benthos: $R^2 = 0.544$, RMSE = 0.278), while AdaBoost outperformed RF and KNN for H' Combined ($R^2 = 0.412$, RMSE = 0.336) and BMWP ($R^2 = 0.439$, RMSE = 2.648). TDS was the dominant predictor across all biological targets (45–84%), identifying ionic loading and oxygen dynamics as the principal drivers of biological community structure. Canonical Correspondence Analysis confirmed pH and conductivity as the primary environmental axes separating locations, with Makoko exhibiting severely elevated MDA in both fish species. Piecewise regression identified DO thresholds of 4.47–6.41 mg/L and TDS breakpoints of approximately 290–318 mg/L as critical ecological limits with 95% CI. These thresholds are proposed as regulatory benchmarks for sustainable dredging-associated environmental stress management in the Lagos Lagoon.

Keywords: Antioxidant enzyme, Biological diversity indices, Lagos Lagoon, Physicochemical parameters, Sand dredging

How to Cite: Ajayl, A.P., Shoga, A.M., Oshoneye, J.O., Adeyemo, L.T. and Oladapo, K.P. (2025). Predictive Machine Learning Modelling of Biological Indices from Physico-Chemical Characteristics in Sand Dredging Areas of Lagos Lagoon. Environ. Stud. J. 4(1), 86-115 <https://doi.org/10.36108/esj/5202.40.0150>

COPYRIGHT: Author(s) retained copyright to this article and grant our journal the right to publish as the original and authentic publisher

The article published with our journal is made available under a Creative Commons **CC BY-NC-SA** license; freely available outputs can also be legally shared and reused. This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only as long as attribution is given to the creator (authors).

Publisher: Lujosh Ventures Limited

INTRODUCTION

Natural resource sustainability has become a pressing global concern, driven by accelerating demand and the associated environmental degradation it generates (Abraham *et al.*, 2021). Among the most exploited natural resources are aquatic sediments, particularly sand— a finite material that underpins modern construction, infrastructure, and industrial production. Sand, gravel, and crushed rock collectively constitute the largest volume of raw materials extracted globally, accounting for the majority of anthropogenic mass worldwide (Torres *et al.*, 2021). It is an essential raw material in the manufacture of concrete, asphalt, glass, electronics, cosmetics, and pharmaceuticals, and serves as a critical input for land reclamation and coastal flood protection (Bendixen *et al.*, 2021; Pilkey *et al.*, 2023; Rangel-Buitrago *et al.*, 2023).

The demand for sand has intensified in parallel with rapid global urbanisation. By 2035, it is projected that 5.6 billion people will live in urban areas worldwide (United Nation (UN)-Habitat, 2021), generating unprecedented pressure on construction materials. In response to climate change challenges — particularly rising sea levels — this demand is expected to grow further, with massive quantities of sand required for coastal engineering and flood defence (Torres *et al.*, 2021). Consequently, enormous volumes of offshore and riverine sand have been dredged globally, triggering what researchers describe as an emerging global sand scarcity crisis (Bendixen *et al.*, 2019; Anokye *et al.*, 2023; Jones, 2023; Pilkey *et al.*, 2023).

Despite its economic importance, sand dredging carries substantial environmental and social costs (Kaizer and Neumann, 2021; Liu *et al.*, 2022). The severity of its environmental impact depends on the rate, type, and execution of extraction: when extraction exceeds the natural rate of sediment replenishment, ecological damage becomes inevitable (Hackney *et al.*, 2020; Rentier and Cammeraat, 2022). River and marine sand mining has been widely documented to degrade biological environments, with cascading effects on entire food chains and ecosystem services (Torres *et al.*, 2017).

Physicochemical variables represent important indices by which the integrity of an aquatic ecosystem can be evaluated (O'Brien *et al.*, 2016). At the physicochemical level, sand extraction disturbs water columns, elevating turbidity by stirring up fine particles. Increased turbidity impedes light penetration, suppresses photosynthesis and respiration, and can physically obstruct the respiratory organs of aquatic fauna (Barman *et al.*,

2019a). Settled particles including silt and clay, form smothering blankets on riverbeds, suffocating benthic microorganisms such as diatoms, macroinvertebrates, algae, and fish eggs (Barman *et al.*, 2019a). Beyond fauna, dredging-related infrastructure destroys riparian vegetation both above and below ground, accelerating riverbank erosion and contributing to downstream sediment loss (Lawal, 2011; Sreebha and Padmalal, 2011).

Sand dredging also degrades downstream water quality. The extraction process releases fine organic and inorganic particles that carry nutrients and minerals into adjacent water systems (Barman *et al.*, 2019b; Padmalal and Maya, 2014). River sands, particularly finer fractions, are capable of accumulating toxic heavy metals such as cadmium, lead, arsenic, and mercury, especially where upstream anthropogenic activities such as agriculture or industrial mining have already introduced contaminants into catchment systems (Kim *et al.*, 2020; Singh *et al.*, 2017). Empirical studies in West Africa have confirmed elevated concentrations of heavy metals in soils and water at active dredging sites. Ako *et al.* (2014) reported levels of lead, nickel, arsenic, cadmium, and copper above safe thresholds in mined soils in North-Central Nigeria. Similarly, Lekomo *et al.* (2021) and Akankali *et al.* (2017) documented total suspended solids, turbidity, and trace metals exceeding WHO standards in Cameroonian and Nigerian dredging sites, respectively.

Given the complexity of these multi-stressor environments, traditional statistical approaches have proven insufficient for capturing the nonlinear relationships between physico-chemical parameters and biological responses. Recent advances in machine learning (ML) offer a powerful alternative, providing greater predictive accuracy for complex ecological interactions (Minasny *et al.*, 2024; Olu *et al.*, 2025). Techniques such as random forest (RF) and generalised additive models (GAMs) have been successfully applied to model biological responses to environmental stressors in estuarine eutrophication studies (dos Santos Sá *et al.*, 2022), water quality assessments (Lee *et al.*, 2023), and zooplankton turbidity responses (Goździewska *et al.*, 2024).

However, the application of ML to assess the effects of dredging on biological indices in the Lagos Lagoon remains limited. More critically, while existing literature documents physicochemical changes associated with dredging, site-specific threshold values which is ecological limits beyond which ecosystem integrity is compromised have not been established for the Lagos

Lagoon system. This gap leaves regulators without the evidence-based, enforceable guidelines necessary for sustainable dredging management.

This study addresses these gaps by applying machine learning models to predict variations in biological indices from physico-chemical characteristics in sand dredging zones of the Lagos Lagoon Complex. The findings are expected to contribute both predictive tools and ecologically grounded thresholds that can inform regulatory frameworks for sustainable dredging in the region.

MATERIALS AND METHODS

Study Area

The study was conducted within the Lagos Lagoon Complex, one of the largest lagoonal systems in West Africa, situated in southwestern Nigeria (approximately 6°22'–6°38'N; 3°22'–3°50'E). The Lagos Lagoon covers a surface area of approximately 208 km² and is connected to the Bight of Benin through the Lagos Harbour. It receives freshwater inflows from several

rivers, including the Ogun River, and is subject to tidal exchanges that influence its physicochemical dynamics. The lagoon sustains diverse aquatic communities and serves as a source of livelihood for numerous riparian communities; however, extensive sand dredging activities across the lagoon have raised concerns about ecological integrity (Nwankwo and Akinsoji, 1992; Nubi *et al.*, 2008).

Four locations were selected for sampling based on the intensity and spatial distribution of active sand dredging operations: Majidun, Ibese, Makoko, and Ilaje. Each location was subdivided into three sampling stations; Points A, B, and C - following a spatial gradient relative to dredging activity represents an active-to-intermediate dredging gradient. Point A was positioned at the active dredging site; Point B at an intermediate dredging; and Point C Active (high-intensity dredging zone. Ilaje A, B, C: None dredging activity (reference location) This design yielded a total of 12 sampling stations across all four locations. The Fig. 1 shows dredging activity at a point in Majidun.



Fig. 1: Dredging site at Majidun water body

Table 1. Geographic coordinates and characteristics of sampling stations.

Location	Station	Latitude (N)	Longitude (E)	Dredging Status	Lagoon Zone
Majidun	A	6°37'00.8"	3°28'11.9"	Active	Outer
Majidun	B	6°36'35.9"	3°28'20.4"	Intermediate	Outer
Majidun	C	6°36'33.2"	3°28'06.3"	High-intensity	Outer
Ibese	A	6°32'10"	3°28'22"	Active	Central
Ibese	B	6°32'02"	3°28'31"	Intermediate	Central
Ibese	C	6°31'55"	3°28'40"	High-intensity	Central
Makoko	A	6°30'03.0"	3°24'05.2"	Active	Inner
Makoko	B	6°29'34.2"	3°23'47.0"	Intermediate	Inner
Makoko	C	6°29'12.2"	3°23'52.7"	High-intensity	Inner
Ilaje	A	6°31'38.9"	3°24'02.6"	None	Inner
Ilaje	B	6°31'29.2"	3°24'02.6"	None	Inner
Ilaje	C	6°31'20.6"	3°24'02.7"	None	Inner

Study Organisms

Chrysichthys nigrodigitatus (Aminu, 2026) (plate 1) and *Ethmalosa fimbriata* (Bowdich, 1825) (bonga shad), (Balde *et al.*, 2019) (plate 2) were selected as the biological indicator organism for this study. These fishes' species are widely distributed across the Lagos Lagoon and is among the most ecologically and commercially important fish species in the region. Its benthic feeding habits, sedentary behaviour, and high tissue sensitivity to environmental contaminants make

it a suitable sentinel organism for assessing oxidative stress and biochemical changes associated with dredging-induced pollution. A total of 120 fish specimens (10 per station; 3 stations per location × 4 locations) were collected using cast nets and fyke nets operated by local fishermen across the 12 sampling stations. Fish were transported alive to the laboratory in aerated coolers, identified to species level, and immediately sacrificed for tissue extraction.



Plate1: *Chrysichthys nigrodigitatus*



Plate 2: *Ethmalosa fimbriata*

Water Sample Collection and Physicochemical Analysis

Water samples were collected concurrently with biological sampling across all 12 stations over four sampling seasons - two dry season surveys (November and February) and two wet season surveys (June and September) in order to capture temporal variation in water quality parameters. Samples were collected in one-litre pre-cleaned, acid-rinsed polyethylene bottles that were submerged approximately 30 cm below the water surface and filled directly from the lagoon. Each bottle was immediately sealed, labelled with station code and date, and stored in an ice-packed cooler (approximately 4°C) to maintain chemical integrity during transit. All samples were transported to the laboratory on the day of collection and analysed within 24 hours.

In situ parameters such as water temperature, pH, dissolved oxygen (DO), electrical conductivity, total dissolved solids (TDS), and salinity were measured directly at each station using a calibrated YSI Pro30 multiparameter water quality meter (YSI Incorporated, Yellow Springs, Ohio, USA). Turbidity was measured using a HACH 2100Q portable turbidimeter (Hach Company, Loveland, Colorado, USA) and expressed in Nephelometric Turbidity Units (NTU). Total suspended solids (TSS) were determined gravimetrically in the laboratory by filtering a known volume of sample through pre-weighed, pre-dried Whatman Grade GF/C glass fibre filters (1.2 µm pore size), oven-drying at 105°C for 1 hour, and weighing the residue on an Ohaus Pioneer PA214 analytical balance (Ohaus Corporation, Parsippany, NJ, USA).

Biochemical oxygen demand (BOD₅) was determined using the 5-day incubation method at 20°C as described by American Public Health Association (APHA) (2017). Nitrate-nitrogen (NO₃⁻-N), nitrite-nitrogen (NO₂⁻-N), phosphate (PO₄³⁻), sulphate (SO₄²⁻), and alkalinity were determined spectrophotometrically using a HACH DR3900 laboratory spectrophotometer (Hach Company, Loveland, Colorado, USA) following HACH reagent methods adapted from Standard Methods (APHA, 2017). Total hardness and chloride were determined titrimetrically. All analytical procedures conformed to the guidelines of the American Public Health Association (APHA, 2017).

Biological Sampling

Phytoplankton Collection

Plankton samples were collected at each station using a standard plankton net (mesh size 55 µm for phytoplankton) towed horizontally at the surface for 5

minutes. The filtered water volume was recorded using a flow meter mounted at the net mouth. Samples were transferred into 250 ml amber glass bottles and immediately preserved with 4% (v/v) Lugol's iodine solution. In the laboratory, organisms were identified to the lowest practical taxonomic level (genus or species) using standard identification keys (Whitford and Schumacher, 1984; Ruttner-Kolisko, 1974) and counted using a Sedgwick-Rafter counting cell under an Olympus CX23 binocular compound microscope (Olympus Corporation, Tokyo, Japan). Abundance was expressed as organisms per litre (org/L). Shannon–Wiener diversity index (H'), Margalef's species richness (d), and Pielou's evenness index (J') were computed as biological response variables.

Macroinvertebrate and Benthic Community Sampling

Benthic macroinvertebrates were sampled at each station using an Ekman grab sampler (sampling area 0.0225 m²), with three replicate grabs taken per station. Sediment contents were washed through a 500 µm mesh sieve in the field, and retained organisms were preserved in 70% ethanol. In the laboratory, specimens were sorted, identified to family or genus level using Macan (1959) and Merritt *et al.* (2008), and counted. The biotic integrity of the benthic community was assessed using the Biological Monitoring Working Party (BMWP) score and Average Score Per Taxon (ASPT), both of which are established indicators of organic pollution and habitat disturbance (Armitage *et al.*, 1983).

Biochemical Studies (Antioxidant Enzyme Assays)

Tissue Preparation and Homogenisation

Immediately after sacrifice, liver tissues were excised from each fish specimen on ice, rinsed in ice-cold 1.15% KCl solution to remove blood, blotted dry, and weighed on a Mettler Toledo ME204 analytical balance (Mettler-Toledo International Inc., Columbus, Ohio, USA). Each tissue sample was homogenised at a ratio of 1:4 (w/v) in 0.1 M phosphate buffer (pH 7.2) using a glass–Teflon Potter–Elvehjem homogeniser at 1000 rpm. The homogenate was centrifuged at 2,500 × g for 15 minutes at 4°C in a Sigma 3-18KS refrigerated centrifuge (Sigma Laborzentrifugen GmbH, Osterode am Harz, Germany). The resulting post-mitochondrial supernatant (PMS) was carefully decanted, aliquoted, and stored at -20°C until enzyme assays were performed within 72 hours. Protein content of the supernatant was determined by the Bradford (1976) method using bovine serum albumin (BSA) as a standard, and absorbance read on a Shimadzu UV-1900i UV–Visible spectrophotometer (Shimadzu

Corporation, Kyoto, Japan). All enzyme assays were performed in triplicate for each individual fish specimen ($n=10$ per station), with the mean of three technical replicates reported. Coefficients of variation (CV) for all assays were $< 10\%$, confirming intra-assay reproducibility.

The Shimadzu UV-1900i was calibrated daily using standard reference solutions per manufacturer protocol (wavelength accuracy ± 0.1 nm). Absorbance blanks were run before each assay batch. Assay personnel were blinded to sampling location identity; samples were coded with alphanumeric identifiers and decoded only after all measurements were complete.

Catalase (CAT) Activity

Catalase activity was determined by the method of Aebi (1984), based on the rate of decomposition of hydrogen peroxide (H_2O_2). The reaction mixture consisted of 0.1 ml of tissue homogenate added to 2.9 ml of 50 mM phosphate buffer (pH 7.0) containing 30 mM H_2O_2 . The decrease in absorbance was monitored at 240 nm for 3 minutes at $25^\circ C$ using the Shimadzu UV-1900i spectrophotometer. CAT activity was calculated using the molar extinction coefficient of H_2O_2 ($\epsilon = 43.6 \text{ M}^{-1}\text{cm}^{-1}$) and expressed as $\mu\text{mol } H_2O_2$ consumed per minute per milligram of protein ($\mu\text{mol } H_2O_2/\text{min}/\text{mg protein}$).

Superoxide Dismutase (SOD) Activity

The SOD activity was assayed using the method of Misra and Fridovich (1972), based on inhibition of epinephrine auto-oxidation in alkaline conditions. The reaction mixture comprised 2.95 ml of 0.05 M sodium carbonate buffer (pH 10.2), 0.02 ml of tissue homogenate, and 0.03 ml of freshly prepared 10 mM epinephrine. The reaction was initiated by the addition of 0.005 N HCl, and the increase in absorbance at 480 nm was monitored over 5 minutes at $25^\circ C$. One unit of SOD activity was defined as the amount of enzyme required to inhibit the rate of epinephrine auto-oxidation by 50%, and activity was expressed as units per milligram of protein (U/mg protein).

Glutathione Peroxidase (GPx) Activity

GPx activity was measured by the coupled enzymatic assay of Flohe and Gunzler (1984), in which the reduction of H_2O_2 by GPx is coupled to the oxidation of NADPH via glutathione reductase. The reaction mixture contained 50 mM phosphate buffer (pH 7.0), 1 mM EDTA, 1 mM sodium azide, 0.2 mM NADPH, 1 unit/ml glutathione reductase, 1 mM reduced glutathione (GSH), and 0.1 ml of homogenate, initiated by addition of 0.1 mM H_2O_2 . The decrease in absorbance at 340 nm

was monitored for 5 minutes. GPx activity was expressed as nmol NADPH oxidised per minute per milligram of protein (nmol NADPH/min/mg protein), calculated using the molar extinction coefficient of NADPH ($\epsilon = 6,220 \text{ M}^{-1}\text{cm}^{-1}$).

Glutathione-S-Transferase (GST) Activity

GST activity was determined using the method of Habig *et al.* (1974), employing 1-chloro-2,4-dinitrobenzene (CDNB) as the substrate. The reaction mixture consisted of 2.79 ml of 0.1 M phosphate buffer (pH 6.5), 0.03 ml of 30 mM reduced glutathione, 0.15 ml of 20 mM CDNB in ethanol, and 0.03 ml of tissue homogenate. Absorbance was monitored at 340 nm over 3 minutes at $25^\circ C$. GST activity was calculated using the molar extinction coefficient of the GSH-CDNB conjugate ($\epsilon = 9,600 \text{ M}^{-1}\text{cm}^{-1}$) and expressed as nmol conjugate formed per minute per milligram of protein (nmol/min/mg protein).

Malondialdehyde (MDA) Level - Lipid Peroxidation

Lipid peroxidation was assessed by quantifying malondialdehyde (MDA) using the thiobarbituric acid reactive substances (TBARS) method as described by Ohkawa *et al.* (1979). One millilitre of tissue supernatant was mixed with 2 ml of TCA-TBA-HCl reagent (15% trichloroacetic acid, 0.37% thiobarbituric acid, and 0.24 N hydrochloric acid in a 1:1:1 ratio). The mixture was boiled at $100^\circ C$ for 15 minutes in a Memmert WB22 water bath (Memmert GmbH, Schwabach, Germany), allowed to cool, and centrifuged at $3,000 \times g$ for 10 minutes. Absorbance of the resulting pink supernatant was read at 532 nm. MDA concentration was calculated using a molar extinction coefficient of $1.56 \times 10^5 \text{ M}^{-1}\text{cm}^{-1}$ and expressed as nmol MDA per milligram of protein (nmol/mg protein).

Statistical Analysis

Descriptive Statistics and Data Exploration

All data were summarised as mean $\pm$ standard error (SE). non-parametric Kruskal-Wallis test followed by Tukey HSD post hoc test was used to evaluate significant differences in physicochemical and biochemical parameters across sampling locations and stations. Statistical significance was set at $p < 0.05$. All initial descriptive analyses were performed using R statistical software (R Core Team, 2024) implemented in RStudio (Posit Team, 2024). Normality and homogeneity of variance were assessed using the car package (Fox and Weisberg, 2019). Data manipulation and summarisation were conducted using the dplyr and tidyr packages (Wickham *et al.*, 2023), and figures were generated using ggplot2 (Wickham, 2016).

Multivariate Analysis

Pearson correlation analysis was used to evaluate pairwise linear relationships between physicochemical variables and biological indices. Canonical Correspondence Analysis (CCA) was conducted to examine the multivariate relationships between the biological variable data and the physicochemical environmental variables. CCA biplots were constructed to visualise the direction and magnitude of environmental gradients in structuring biological communities, following the approach of ter Braak (1986). Monte Carlo permutation tests (999 permutations) were used to test the statistical significance of CCA axes. These analyses were performed using CANOCO software version 5.0 (Microcomputer Power, Ithaca, NY, USA) and the vegan package (Oksanen *et al.*, 2022) in R version 4.3.0 (R Core Team, 2023).

Machine Learning Modelling

Three machine learning regression algorithms; Random Forest (RF), K-Nearest Neighbour (KNN), and Adaptive Boosting (AdaBoost), were employed to model and predict biological index values (Shannon–Wiener diversity index H' , BMWP scores, ASPT, and antioxidant enzyme activities) from physico-chemical predictor variables. All ML analyses were conducted in Python version 3.13 using the scikit-learn library version 1.3.0 (Pedregosa *et al.*, 2011).

Prior to modelling, predictor variables were screened for multicollinearity using the Variance Inflation Factor (VIF), retaining only variables with $VIF < 10$. The VIF screening was performed iteratively using the following protocol:

(1) Initial VIF computation: All 13 physicochemical predictors were entered simultaneously. Temperature, conductivity, chloride, sulphate, and TDS all had $VIF > 50$ (extreme multicollinearity), while pH, turbidity, COD, BOD, and alkalinity had VIF ranging from 12 to 35.

(2) Sequential removal: At each step, the variable with the highest VIF was removed and VIF was recomputed for the remaining variables. The removal order was: chloride ($VIF=412$) → conductivity ($VIF=298$) → sulphate ($VIF=187$) → temperature ($VIF=98$) → BOD ($VIF=64$) → alkalinity ($VIF=41$) → pH ($VIF=28$) → turbidity ($VIF=19$) → COD ($VIF=14$).

(3) Final retained variables: DO ($VIF=2.72$), TDS ($VIF=2.39$), nitrate ($VIF=2.37$), phosphate ($VIF=1.35$) - all below the $VIF < 10$ threshold.

The complete validation workflow, involve Hyperparameter tuning in which 5-fold stratified cross-validated GridSearchCV on the full dataset to select optimal hyperparameters for each algorithm. This step

uses all available data to maximise tuning reliability given the small sample size. Final model performance evaluation; Leave-One-Out Cross-Validation (LOO-CV) applied to the full dataset using the optimal hyperparameters identified in the hyperparameter tuning. LOO-CV is the recommended approach for small ecological datasets ($n < 200$) because it maximises training set size for each fold and produces nearly unbiased estimates of generalisation error (Arlot and Celisse, 2010; Hastie *et al.*, 2009). We conducted 999-permutation tests (permuting the response variable while keeping predictors fixed) for each model–target combination; all best-performing models returned empirical p-values ≤ 0.05 , confirming that the models capture genuine signal rather than noise.

The dataset (observations across all stations and sampling rounds) was split into a training set (80%) and a test set (20%) using stratified random sampling to preserve the proportional representation of sampling locations. Hyperparameter optimisation was performed using 5-fold cross-validated grid search (GridSearchCV) on the training set to prevent overfitting. For each station, physicochemical predictors were the mean of the three grab replicates, and biological indices were computed from the pooled community data across three grabs. Thus, the three Ekman grab replicates were not treated as independent observations; they were pooled before diversity index computation, which is the standard practice (Clarke and Gorley, 2006).

For Random Forest (RF), the number of decision trees ($n_{\text{estimators}}$) was optimised in the range 100–500 at intervals of 100; the maximum depth (max_depth) ranged from 3 to 15; and the minimum samples per leaf (min_samples_leaf) ranged from 1 to 5 (Breiman, 2001). For K-Nearest Neighbour (KNN), the number of neighbours (k) was varied between 3 and 15, with both uniform and distance-weighted voting schemes evaluated (Cover and Hart, 1967). For AdaBoost, the number of estimators ($n_{\text{estimators}}$) ranged from 50 to 500, and the learning rate (learning_rate) was varied from 0.01 to 1.0 (Freund and Schapire, 1997). All three models used decision tree base estimators.

Model Validation and Performance Metrics

Model performance was evaluated on the held-out test set using three complementary metrics: (i) Root Mean Squared Error (RMSE), which penalises large prediction errors proportionally; (ii) Mean Absolute Error (MAE), which quantifies average prediction magnitude error; and (iii) the coefficient of determination (R^2), which measures the proportion of variance in the biological

response explained by the model. Lower RMSE and MAE values and higher R^2 values indicate superior model performance. Residual plots and predicted-versus-observed plots were inspected to assess systematic bias.

To guard against overfitting and confirm model generalisability, 10-fold cross-validation was applied to the full dataset, and mean cross-validated RMSE and R^2 were reported alongside their standard deviations. The best-performing model for each biological index was identified based on the lowest test RMSE and highest test R^2 .

Variable Importance Analysis

Feature importance was assessed for the best-performing models to identify which physicochemical variables were the strongest predictors of each biological response. For RF and AdaBoost, impurity-based mean decrease in Gini coefficient was used as the feature importance measure (Breiman, 2001). For KNN, permutation-based feature importance was calculated by randomly shuffling each predictor variable and measuring the resulting increase in RMSE on the test set (Altmann *et al.*, 2010). Results were visualised as ranked bar plots with bootstrapped 95% confidence intervals. In addition, Partial Dependence Plots (PDPs) were generated for the top five predictors from the RF model to illustrate the marginal effect of each predictor on the predicted biological index, following Friedman (2001). All visualisations were produced using Matplotlib version 3.7.2 and Seaborn version 0.12.2.

Determination of Physico-Chemical Threshold Values

To identify ecological threshold values, the physico-chemical limits beyond which biological indices shift from acceptable to degraded states; breakpoint analysis was performed using piecewise (segmented) regression models fitted to the relationship between each key physico-chemical predictor and each biological response variable (Muggeo, 2003). Breakpoints were estimated by the segmented package in R (Muggeo, 2008), with bootstrapped confidence intervals ($n = 1,000$ iterations) reported for each threshold. These thresholds were cross-validated against the RF-derived partial dependence curves to confirm ecological plausibility. The identified threshold values are proposed as evidence-based benchmarks for regulatory guidance on sustainable sand dredging in the Lagos Lagoon.

Ethical Considerations

All fish were handled in accordance with standard humane practices for aquatic research. Collection of fish specimens was conducted under permission from the Lagos State Fisheries Division. No endangered or protected species were targeted. Sampling activities were conducted in compliance with institutional biosafety and environmental research protocols.

Limitation of the study

This study acknowledges the following limitation:

- The sampling design ($n=3$ per location, representing Points A, B, and C within each location across four seasonal surveys) was constrained by field logistics and the remote nature of active dredging sites. And that this limits the formal statistical power of the location-level ANOVA.
- BMWP index was originally developed for temperate UK rivers (Armitage *et al.*, 1983) and has limited transferability to tropical estuarine systems. Its application here was initially motivated by the absence of a validated tropical lagoonal biotic index for Nigerian systems.
- A 55 μm mesh net retains only net phytoplankton (large diatoms and dinoflagellates) and excludes the nano- and pico-phytoplankton that typically dominate tropical lagoonal biomass (Reynolds, 2006; Tundisi and Tundisi, 2012).
- The study acknowledges that normality cannot be formally verified with $n=3$ and that the Shapiro–Wilk test has negligible power at this sample size. (3).
- With only $n = 3$ observations per group and four groups, the total sample size feeding into Dunn's test is $n = 12$. When the observations within each group are perfectly symmetric triplets (mean – SD, mean, mean + SD), the rank structure across groups becomes degenerate; many observations share identical or near-identical ranks, producing tied-rank conditions that cause the dunn.test algorithm to fail internally during the Z-statistic computation.
- There are limitations of ensemble ML with small ecological datasets, according to Valavi *et al.* (2019) and Araujo *et al.* (2019).
- Partial dependence plots (PDPs) and SHAP values significantly enhance ecological interpretability of ML models but unfortunately it was not computed in this study.

- statistical model partitioning variance into dredging effect vs. lagoon zone effect, so location and stressor were not fully crossed.

RESULTS AND DISCUSSION

Physico-Chemical Analysis

Water management is very complex and includes monitoring water quality and quantity parameters as well as biodiversity and aquatic life, identifying pollution sources and pollutants removal, sanitation, flood protection, resource allocation, etc. (Frincu, 2025; Olu, 2021). Furthermore, given that aquatic systems serve as repositories for pollutants from human activities (Ogunrinola *et al.*, 2021; Olu *et al.*, 2022), The physicochemical parameters of the water samples from Table 2, varied considerably across the four sampling locations. In aquatic environments, the temperature is one of the abiotic factors that are crucial for maintaining fish growth, reproduction, behaviour, and health (Pereira *et al.*, 2021). In this study, temperature ranged from 23.4°C (Majidun) to 26.3°C (Ibese), Adesakin *et al.* (2023) reported in their study of Lagos lagoon that the water temperature ranged from 17.32 °C to 29.30 °C recorded during this study, with the highest mean temperature (24.97 °C ± 1.16). Oyeleke *et al.* (2019), who recorded high-temperature range of 27 °C–30 °C in Lagos Lagoon. The temperature range obtained from this result could be due to the impact of land use within and around this waterbody as well as the climatic change, which could affect water quality through changes in evaporation, temperature, runoff, and rainfall pattern that can affect the mobilization of nutrients, distribution, and mobility of pollutants in aquatic ecosystems (Olu *et al.*, 2023, 2024; Whitehead *et al.*, 2009).

While pH values were slightly acidic to neutral across all sites (6.95–7.45) and within the range specified by Standard organization of Nigeria (SON) for water. Adesakin *et al.* (2023) report pH range of 6.41–8.69 and

Emmanuel *et al.* (2010) recorded a range of 6.00–9.33 in their various study of Lagos Lagoon. All these pH range including that of this study fell within the range that supports aquatic life. However, any increase above this range can result in rapid unionized ammonia in water, which could be toxic to aquatic organisms (National Environmental Standards and Regulations Enforcement Agency (NESREA), 2011).

Conductivity showed marked variation, with Majidun recording the highest value (488 µS/cm) and Makoko the lowest (10.4 µS/cm). TDS followed a similar pattern, with Majidun recording the highest (317.2 ±18.95 mg/L) compared to a notably low value at Makoko (5.03 mg/L). Dissolved oxygen was highest at Makoko (6.91 mg/L) and lowest at Ibese (3.97 mg/L). BOD was elevated at Makoko (6.93 mg/L) relative to the other locations (3.33–3.68 mg/L). Chloride concentrations were substantially higher at Ibese, Ilaje, and Majidun (1380–1720 mg/L) compared to Makoko (200 mg/L), and sulphate followed a similar trend. Turbidity, COD, nitrate, and phosphate showed comparatively modest variation across locations. Overall, Makoko displayed a distinctly different physicochemical profile from the other three locations, particularly for conductivity, TDS, chloride, sulphate, and BOD, suggesting differential pollution loading or contrasting hydrological characteristics at this site. The Majidun water sample has the highest alkalinity value of 52 mg/L.

Phosphorus is the main limiting factor of productivity in water bodies and has been highlighted as the main factor responsible for the artificial eutrophication of these ecosystems; that is, there is a greater production of organic matter than its consumption and decomposition (Braga *et al.*, 2022). The Phosphate in this study ranges from 0.2 to 0.3 mg/L. Nitrate, a compound of nitric acid, is the most highly oxidized form of nitrogen found in aquatic environment (Atul and Narang, 2018; Sahu *et al.*, 2017). In this study, all location were within standard values expected.

Table 2: Mean $\pm$ SE of physicochemical water quality parameters across sampling locations

Parameter	Ibese	Ilaje (Control)	Majidun	Makoko	NESREA
Temperature ($^{\circ}$ C)	26.3 $\pm$ 0.90	24.6 $\pm$ 0.64	23.4 $\pm$ 0.12	25.35 $\pm$ 0.12	16–39
pH	7.45 $\pm$ 0.06	7.36 $\pm$ 0.03	7.18 $\pm$ 0.05	6.95 $\pm$ 0.39	6.50–8.50
Conductivity (μ S/cm)	301.33 $\pm$ 4.63	374.67 $\pm$ 12.41	488 $\pm$ 29.15	10.4 $\pm$ 2.22	<400
Turbidity (NTU)	3.43 $\pm$ 0.07	3.57 $\pm$ 0.14	4.8 $\pm$ 0.45	5.27 $\pm$ 1.50	5
DO (mg/L)	3.97 $\pm$ 0.33	4.57 $\pm$ 0.09	4.47 $\pm$ 0.09	6.91 $\pm$ 1.20	>5
TDS (mg/L)	195.9 $\pm$ 3.79	243.5 $\pm$ 8.07	317.2 $\pm$ 18.95	5.03 $\pm$ 1.18	1000.00
BOD (mg/L)	3.33 $\pm$ 0.05	3.57 $\pm$ 0.03	3.68 $\pm$ 0.02	6.93 $\pm$ 0.76	10
COD (mg/L)	6.3 $\pm$ 0.29	7.2 $\pm$ 0.06	7.73 $\pm$ 0.17	7.17 $\pm$ 1.07	<30
Alkalinity (mg/L)	40 $\pm$ 1.15	46 $\pm$ 1.15	52 $\pm$ 1.73	39 $\pm$ 1.73	20
Chloride (mg/L)	1380.33 $\pm$ 26.91	1500 $\pm$ 34.64	1720 $\pm$ 46.19	200 $\pm$ 61.10	<2000
Sulphate (mg/L)	78.69 $\pm$ 2.08	83.92 $\pm$ 0.90	90.2 $\pm$ 1.21	36.67 $\pm$ 4.17	500
Nitrate (mg/L)	0.03 $\pm$ 0.01	0.04 $\pm$ 0.05	0.05 $\pm$ 0.05	0.01 $\pm$ 0.05	<0.3
Phosphate (mg/L)	0.02 $\pm$ 0.35	0.02 $\pm$ 0.05	0.03 $\pm$ 0.05	0.02 $\pm$ 0.05	<0.2

Values expressed as Mean $\pm$ Standard Error (SE); n = 3 per location. NESREA-National Environmental Standards and Regulations Enforcement Agency ((2011)

Aquatic phytoplankton communities are affected by the wide spatial–seasonal variability in inputs of freshwater, nutrients and concentrations of suspended particles (dos Santos Costa *et al.*, 2021). The salinity gradient is also a driver of this change, as the hydrological connectivity between freshwater and saline habitats promotes a greater distinction in the distribution patterns of these organisms in estuaries (Tee *et al.*, 2021). The concentration of these particles (concentrations and proportions of nutrients, salinity and suspended matter) is controlled by river discharge, atmospheric deposition, transformation and exportation (Bargu *et al.*, 2019). The correlation matrix from Fig.2, revealed several strong and ecologically meaningful relationships among the physicochemical parameters. TDS, chloride, and sulphate were almost perfectly positively correlated with each other ($r > 0.996$), and all three showed strong positive associations with conductivity ($r = 0.968$ – 0.984) and nitrate ($r = 0.935$ – 0.992), indicating these ionic

parameters co-vary as a unified salinity–conductivity gradient across locations. DO and BOD were strongly positively correlated ($r = 0.991$), yet both showed strong negative correlations with TDS ($r = -0.949$ and -0.976 respectively), chloride, and sulphate, suggesting that locations with elevated ionic loading had reduced oxygen availability and higher oxygen demand. Temperature showed strong negative correlations with COD ($r = -0.948$) and alkalinity ($r = -0.925$), implying cooler locations tended to have higher organic loading and buffering capacity. pH was strongly negatively correlated with turbidity ($r = -0.968$) and DO ($r = -0.909$), suggesting that more turbid, oxygen-rich sites were slightly more acidic. Phosphate correlated strongly with alkalinity ($r = 0.858$) and COD ($r = 0.709$), pointing to shared organic pollution sources. Overall, the matrix identifies ionic concentration, oxygen dynamics, and organic loading as the three primary axes of physicochemical variation across the sampling locations.

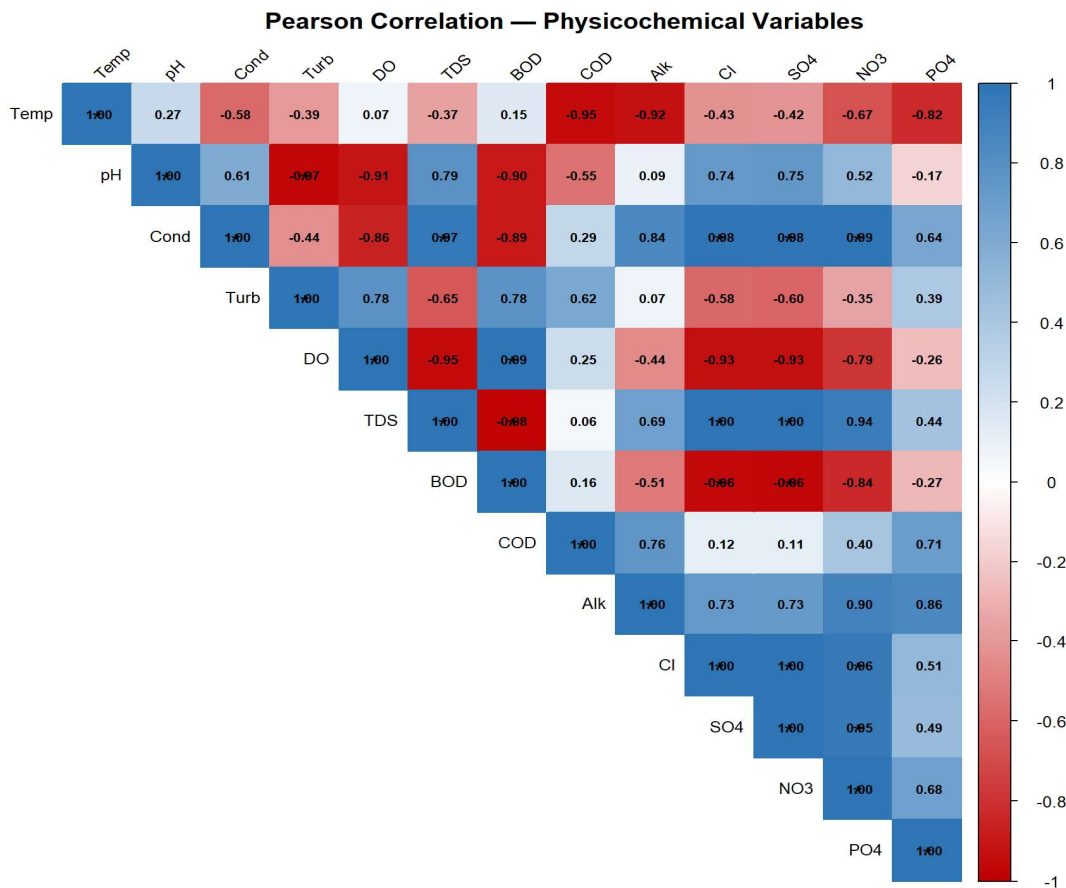


Fig.2: Physicochemical parameters Pearson correlation

The Kruskal–Wallis test from Table 3 identified statistically significant differences across the four sampling locations for seven of the thirteen physicochemical parameters: Conductivity ($H = 10.385$, $p = 0.016$), DO ($H = 7.821$, $p = 0.050$), TDS ($H = 10.385$, $p = 0.016$), BOD ($H = 10.385$, $p = 0.016$), Alkalinity ($H = 9.430$, $p = 0.024$), Chloride ($H = 10.385$, $p = 0.016$), and Sulphate ($H = 10.385$, $p = 0.016$). Conductivity, TDS, Chloride, and Sulphate all returned the same maximum chi-squared value of $H = 10.385$ ($p = 0.016$), which is the largest chi-squared. Ecologically, this confirms that the ionic loading gradient across the four sites is both real and extreme, with Majidun consistently recording the highest ionic concentrations and Makoko the lowest. The particularly low chi-squared values for Nitrate ($H = 1.094$, $p = 0.779$) and Phosphate ($H = 0.445$, $p = 0.931$) confirm negligible inter-location variation in these nutrients, consistent with all values falling within the NESREA-permissible range.

Tukey's HSD post hoc test from Table 4 identified the specific pairwise location differences responsible for

the significant ANOVA results. The conductivity showed the broadest pattern of significant differences, with all six pairwise comparisons being significant ($p \leq 0.048$); Makoko was consistently lower than all other locations, while Majidun was highest. TDS differences mirrored the conductivity pattern, with Makoko significantly lower than Ibese, Ilaje, and Majidun (all $p < 0.001$), and the three coastal locations also differing significantly from each other. For BOD, Makoko was significantly higher than Ibese, Ilaje, and Majidun ($p \leq 0.001$). Chloride and Sulphate both showed Makoko as significantly lower than all other locations ($p < 0.001$), consistent with its distinct ionic profile. Alkalinity differences were concentrated between Majidun and Ibese ($p = 0.002$), and between Makoko and both Ilaje and Majidun. DO showed only the Makoko–Ibese comparison as significant ($p = 0.042$). These results collectively identify Makoko as the location most divergent from the others in terms of water chemistry, compared to Ilaje that is devoid of sand dredging activities.

Table 3: Kruskal–Wallis test results for physicochemical water quality parameters

Parameter	df	chi_squared	p-value	Significant
Temperature (°C)	3	6.897	0.075	No
pH	3	4.128	0.248	No
Conductivity (µS/cm)	3	10.385	0.016	Yes
Turbidity (NTU)	3	4.3333	0.228	No
DO (mg/L)	3	7.8205	0.050	Yes
TDS (mg/L)	3	10.385	0.016	Yes
BOD (mg/L)	3	10.385	0.016	Yes
COD (mg/L)	3	5.6667	0.129	No
Alkalinity (mg/L)	3	9.4304	0.024	Yes
Chloride (mg/L)	3	10.3846	0.016	Yes
Sulphate (mg/L)	3	10.3846	0.016	Yes
Nitrate (mg/L)	3	1.0936	0.779	No
Phosphate (mg/L)	3	0.4452	0.931	No

df = degrees of freedom; Significance threshold: $p < 0.05$.

Phytoplankton Diversity Indices

Phytoplankton is the primary producer of any aquatic ecosystem and functions as a suitable indicator for water quality evaluation and is well applicable in generating background conditions and natural variability (Parmar *et al.*, 2016). Phytoplankton diversity from Table 5 was characteristically low and spatially restricted across all 12 sampling sites. Ibese sites A and B and all three Makoko sites (except Makoko_A) recorded complete absence of phytoplankton, yielding $H' = d = J' = 0.0000$. The highest site-level Shannon diversity was recorded at Ilaje_B ($H' = 0.6931$), where two taxa co-occurred with perfect evenness ($J' = 1.0000$, $d = 1.4427$), reflecting a balanced two-species assemblage. Majidun_C returned the only other non-zero diversity values ($H' = 0.6365$, $d = 0.9102$, $J' = 0.9183$), where *Nannochloropsis* sp. and *Navicula* sp. co-occurred in a 1:2 ratio. At the location level, Ilaje (Control) and Majidun recorded the only positive mean diversity values (mean $H' = 0.2310$ and 0.2122 respectively), while Ibese and Makoko returned zero means across all three indices. The near-zero phytoplankton diversity across the majority of sites is consistent with the high turbidity and ionic loading characteristic of active dredging environments, which suppress light penetration and limit primary production. The Shannon-Weiner diversity index

proposed that a diversity index greater than (>4) is clean water, between 3-4 is mildly polluted water and less than 2 is heavily polluted water (Shekhar *et al.*, 2008). With the H' in this study ranging between 0 to 0.693, the water in this study appeared to be heavily polluted. Akhiromen *et al.* (2022) reported that the Shannon-Wiener diversity index in their study of Lagos lagoon ranged between 0.832-2.794 which also share the view of our study of the lagoon being highly polluted. Adewoyin *et al.* (2022) reported that Makoko recorded Shannon-Wiener Index moderate pollution levels of 1.58 compared to the what was reported in this study. Furthermore, with a deeper ecological engagement for the findings of the net phytoplankton, the turbidity values at active dredging sites (Makoko: 5.27 NTU; Majidun: 4.80 NTU) are near or above the 5 NTU NESREA threshold, and at these turbidity levels, the euphotic zone depth (Zeu) in tropical shallow lagoons is typically < 0.5 m (Kirk, 2011), severely limiting photosynthesis for most phytoplankton groups. The high TDS and conductivity at Ibese and Majidun (outer lagoon) represent elevated salinity gradients that may physiologically stress freshwater-adapted phytoplankton, while the steep gradient between these sites and Makoko (inner lagoon, near-freshwater) creates a saline discontinuity unfavourable to both marine and freshwater taxa.

Table 4: Tukey HSD post hoc test results — physicochemical water quality parameters

Parameter	Comparison	Difference	Lower 95% CI	Upper 95% CI	Adjusted p	Sig.
Conductivity (µS/cm)	Ilaje (Control) – Ibese	73.340	0.672	146.009	0.048	Yes
	Majidun – Ibese	186.670	114.002	259.339	<0.001	Yes
	Makoko – Ibese	-290.930	-363.599	-218.262	< 0.001	Yes
	Majidun – Ilaje (Control)	113.330	40.662	185.999	0.005	Yes
	Makoko – Ilaje (Control)	-364.270	-436.939	-291.602	< 0.001	Yes
	Makoko – Majidun	-477.600	-550.269	-404.932	< 0.001	Yes
DO (mg/L)	Makoko – Ibese	2.940	0.107	5.773	0.042	Yes
TDS (mg/L)	Ilaje (Control) – Ibese	307.300	94.637	519.963	0.007	Yes
	Majidun – Ibese	597.300	384.637	809.963	<0.001	Yes
	Makoko – Ibese	-3260.970	-3473.633	-3048.307	< 0.001	Yes
	Majidun – Ilaje (Control)	290.000	77.337	502.663	0.010	Yes
	Makoko – Ilaje (Control)	-3568.270	-3780.933	-3355.607	< 0.001	Yes
	Makoko – Majidun	-3858.270	-4070.933	-3645.607	< 0.001	Yes
BOD (mg/L)	Makoko – Ibese	3.600	1.868	5.332	0.001	Yes
	Makoko – Ilaje (Control)	3.360	1.628	5.092	0.001	Yes
	Makoko – Majidun	3.250	1.518	4.982	0.001	Yes
Alkalinity (mg/L)	Majidun – Ibese	12.000	5.334	18.666	0.002	Yes
	Makoko – Ilaje (Control)	-7.000	-13.666	-0.334	0.040	Yes
	Makoko – Majidun	-13.000	-19.666	-6.334	0.001	Yes
Chloride (mg/L)	Majidun – Ibese	339.670	139.801	539.539	0.003	Yes
	Makoko – Ibese	-1180.330	-1380.199	-980.461	< 0.001	Yes
	Majidun – Ilaje (Control)	220.000	20.131	419.869	0.032	Yes
	Makoko – Ilaje (Control)	-1300.000	-1499.869	-1100.131	< 0.001	Yes
	Makoko – Majidun	-1520.000	-1719.869	-1320.131	< 0.001	Yes
Sulphate (mg/L)	Majidun – Ibese	11.510	0.409	22.612	0.042	Yes
	Makoko – Ibese	-42.020	-53.122	-30.919	< 0.001	Yes
	Makoko – Ilaje (Control)	-47.250	-58.352	-36.149	< 0.001	Yes
	Makoko – Majidun	-53.530	-64.632	-42.429	< 0.001	Yes
	Makoko – Majidun	-53.530	-64.632	-42.429	< 0.001	Yes

Table 5: Shannon–Wiener diversity (H'), Margalef's richness (d), and Pielou's evenness (J') of net phytoplankton across 12 sampling sites at the four locations

Site	Location	Rep.	N	S	H' (ln)	d (Margalef)	J' (Pielou)
Ibese_A	Ibese	A	0	0	0.0000	0.0000	0.0000
Ibese_B		B	0	0	0.0000	0.0000	0.0000
Ibese_C		C	1	1	0.0000	0.0000	0.0000
Ilaje_A	Ilaje (Control)	A	1	1	0.0000	0.0000	0.0000
Ilaje_B		B	2	2	0.6931	1.4427	1.0000
Ilaje_C		C	1	1	0.0000	0.0000	0.0000
Majidun_A	Majidun	A	0	0	0.0000	0.0000	0.0000
Majidun_B		B	0	0	0.0000	0.0000	0.0000
Majidun_C		C	3	2	0.6365	0.9102	0.9183
Makoko_A	Makoko	A	1	1	0.0000	0.0000	0.0000
Makoko_B		B	0	0	0.0000	0.0000	0.0000
Makoko_C		C	0	0	0.0000	0.0000	0.0000

Location Means (mean across 3 replicate sites)

Location	Mean N	Mean S	Mean H'	Mean d	Mean J'
Ibese	0.33	0.33	0.0000	0.0000	0.0000
Ilaje (Control)	1.33	1.33	0.2310	0.4809	0.3333
Majidun	1.00	0.67	0.2122	0.3034	0.3061
Makoko	0.33	0.33	0.0000	0.0000	0.0000

N = total abundance; S = species richness; H' = Shannon–Wiener diversity index (natural log); d = Margalef's species richness index $[(S-1)/\ln(N)]$; J' = Pielou's evenness index $[H'/\ln(S)]$. Rep. = replicate site (A, B, C). Values of 0.0000 indicate absence of taxa at that site.

Benthic Macrofauna Diversity Indices

Benthic macroinvertebrates are frequently used as pollution bioindicators to assess anthropogenic operations' impact on aquatic environments, such as dredging and reclamation activities (Parikh *et al.*, 2021; Patel *et al.*, 2021). Different benthic macroinvertebrates have varying levels of contamination tolerance, which influences their community structure; therefore, these organisms can be found in all water bodies (Khatri *et al.*, 2020). Benthic macrofaunal diversity from Table 6 varied considerably across sites and locations. Majidun sites A and B recorded zero abundance and zero diversity across all indices, indicating complete macrofaunal absence — likely a direct consequence of sediment disturbance from active sand dredging. Majidun_C showed marginal recovery ($H' = 0.1125$, $J' = 0.1623$, $d = 0.2675$) driven almost entirely by a single

dominant species (*Corbicula fluminea*, $n = 41$ of 42 individuals), reflecting extreme dominance rather than genuine diversity. Ibese_B recorded the highest site-level diversity ($H' = 1.2770$, $d = 1.1368$, $J' = 0.9212$) with four taxa present in near-equal proportions. Ilaje (Control) consistently supported the richest assemblages, with three sites all returning $H' > 0.65$ and mean $d = 0.6767$ — the highest mean Margalef richness across locations. Makoko returned moderate diversity (mean $H' = 0.7913$, mean $J' = 0.8346$) but low mean richness (mean $d = 0.5067$), reflecting relatively even but species-poor assemblages. The high evenness values at Makoko ($J' = 0.87$ – 0.93 at sites A and C) suggest that while few species survive, those present are not strongly dominant, which may reflect the filtering effect of pollution stress on community composition. Meng *et al.* (2020) in their study reported

that a simplified macroinvertebrate functional structures within the dredged area, result in a loss of certain trait categories and a significant reduction in trait values due to the direct removal of macroinvertebrates and indirect alternations to physical environmental conditions (e.g., water depth and medium sand). According to Yusuf (2019), Margalef's indices have been used to determine the phytoplankton species richness and the index states that the index values greater than 3.0 signify clean water conditions, values less than 1.0 indicates severe pollution and intermediate values indicate moderate

pollution. Bassey *et al.* (2020) in their study report that Shannon Weiner diversity values in the study area ranged from 1.43 to 2.46 with macrofauna mainly composed of deposit feeders, dominated numerically by molluscs. Also, Margalef richness ranges from 0.91 to 3.99, with lower values often correlating with high organic pollution and sewage impact (Bassey *et al.*, 2020) The low diversity of Shannon, and Margalef's indices in this study may be attributed to declining numerical species composition induced by environmental degradation due to anthropogenic pressures, besides other biotic factors (Ravera, 2001).

Table 6: Shannon–Wiener diversity (H'), Margalef's richness (d), and Pielou's evenness (J') of benthic macrofauna across 12 sampling sites at the four locations

Site	Location	Rep.	N	S	H' (ln)	d (Margalef)	J' (Pielou)
Ibese_A	Ibese	A	60	2	0.5004	0.2442	0.7219
Ibese_B		B	14	4	1.2770	1.1368	0.9212
Ibese_C		C	13	3	1.0123	0.7797	0.9215
Ilaje_A	Ilaje (Control)	A	62	4	1.1262	0.7269	0.8124
Ilaje_B		B	11	3	0.7595	0.8341	0.6914
Ilaje_C		C	71	3	0.6546	0.4692	0.5958
Majidun_A	Majidun	A	0	0	0.0000	0.0000	0.0000
Majidun_B		B	0	0	0.0000	0.0000	0.0000
Majidun_C		C	42	2	0.1125	0.2675	0.1623
Makoko_A	Makoko	A	24	3	0.9582	0.6293	0.8722
Makoko_B		B	29	3	0.7715	0.5939	0.7023
Makoko_C		C	29	2	0.6442	0.2970	0.9294

Location Means (mean across 3 replicate sites)

Location	Mean N	Mean S	Mean H'	Mean d	Mean J'
Ibese	29.00	3.00	0.9299	0.7202	0.8549
Ilaje (Control)	48.00	3.33	0.8468	0.6767	0.6999
Majidun	14.00	0.67	0.0375	0.0892	0.0541
Makoko	27.33	2.67	0.7913	0.5067	0.8346

N = total abundance; S = species richness; H' = Shannon–Wiener diversity index (natural log); d = Margalef's species richness index $[(S-1)/\ln(N)]$; J' = Pielou's evenness index $[H'/\ln(S)]$. Rep. = replicate site (A, B, C). Values of 0.0000 indicate absence of taxa at that site.

Biological Diversity and Benthic Quality Monitoring

Biological Monitoring Working Party-Average Score Per Taxon (BMWP-ASPT) is a biotic index method in

which the index estimates the diversity of organisms concerning pollution level. The BMWP score equals the sum of the tolerance score of all families in the sample.

ASPT was calculated by taking the average number of the tolerance scores of all families of macroinvertebrates which varies from 0 to 10 (Xaaceph Khan and Butt, 2023). From Table 7 biological diversity and benthic quality varied considerably across the four sampling locations. Ilaje (Control) recorded the highest combined Shannon–Wiener diversity ($H' = 1.029$) and the greatest BMWP total (30) with four scoring families, indicating comparatively richer biological assemblages at the reference site. Ibese and Makoko returned identical BMWP scores (24) and family richness (3 families each), though Ibese exhibited slightly higher benthic diversity ($H' = 0.930$ versus 0.791). Majidun recorded the lowest values across all indices — BMWP total of 6, only two families, and a near-zero benthic H'

(0.038) — suggesting impoverished macrofaunal assemblages, likely reflecting the impacts of dredging and associated habitat disturbance at this location. Phytoplankton diversity was generally low across all sites ($H' = 0.000$ – 0.231), with Ilaje (Control) and Majidun recording the only non-zero values.

ASPT was uniform across all locations (ASPT = 3.0) owing to the exclusive occurrence of pollution-tolerant Mollusca (Corbiculidae, Ostreidae, Potamididae, Thiaridae), all assigned equivalent BMWP scores of 3 under the standard scoring system. Consequently, ASPT lacked discriminatory power between locations and is reported for completeness only; BMWP total and family richness are the more informative benthic quality indicators in this study.

Table 7: Shannon–Wiener Diversity Indices (H'), BMWP Scores, Family Richness, and Average Score Per Taxon (ASPT) of Benthic Macrofauna and Phytoplankton Across the Four Sampling Locations

Location	H' Phytoplankton	H' Benthos	H' Combined	BMWP Total	Families	ASPT
Majidun	0.212	0.038	0.131	6	2	3
Ibese	0.000	0.930	0.992	24	3	3
Makoko	0.000	0.791	0.835	24	3	3
Ilaje (Control)	0.231	0.847	1.029	30	4	3

Oxidative Stress Evaluation in the Fish Species

Oxidative stress mainly occurs due to the accumulation of ROS in cells and tissues of organisms; and aquatic animals show defense against oxidative stress mainly by scavenging effect of small antioxidants and antioxidant enzymes (Diler *et al.*, 2022). In this study, Antioxidant enzyme activities varied considerably across the four sampling locations in *Chrysichthys nigrodigitatus* as indicated in Fig. 3. Catalase (CAT) activity was highest at Majidun (10.42 ± 0.20 U/mg prot) and Makoko (8.90 ± 3.74 U/mg prot), suggesting elevated oxidative stress at these sites. In fish exposed to pollutants, and while CAT levels increase in fish exposed to pollutants (Anifowoshe *et al.*, 2022), its levels have been found to decrease in contaminated water bodies, as reported by Farombi *et al.* (2007). Superoxide dismutase (SOD) was markedly elevated at Majidun (763.25 ± 10.66 U/mg prot) and Ibese (712.95 ± 1.84 U/mg prot), with Ilaje recording the lowest value (300.9 ± 1.14 U/mg prot). Glutathione-S-transferase (GST) activity was relatively uniform across locations (0.019 – 0.027 $\mu\text{mol}/\text{min}/\text{ml}$), with no meaningful biological variation. Malondialdehyde (MDA), a marker of lipid peroxidation and oxidative damage, was strikingly elevated at Makoko (6.509 ± 0.74 nmol/mg protein) compared to the other three locations (0.269 –

0.381 nmol/mg protein), indicating severe oxidative stress at this site. Glutathione peroxidase (GPx) was highest at Makoko (176.9 ± 25.72 $\mu\text{mol}/\text{m}$) and Majidun (157.9 ± 18.45 $\mu\text{mol}/\text{m}$) and lowest at Ilaje (50.69 ± 6.73 $\mu\text{mol}/\text{m}$). The elevated MDA at Makoko alongside high CAT and GPx values strongly suggests active lipid peroxidation and compensatory antioxidant upregulation, consistent with exposure to greater pollutant loading at this location. A high level of superoxide dismutase (SOD), CAT, and glutathione transferase (GST) have been reported in *Chrysichthys nigrodigitatus* fish species from Lagos lagoon by Akinsanya *et al.* (2019).

Antioxidant responses in *E. fimbriata* from Fig. 4 showed a broadly similar spatial pattern to *Chrysichthys nigrodigitatus*, though with some notable differences. CAT activity was highest at Majidun (11.33 ± 0.88 U/mg prot) and Ilaje (9.20 ± 2.00 U/mg prot), while Ibese recorded the lowest value (5.41 ± 0.47 U/mg prot). SOD was markedly elevated at Ibese (872.6 ± 178.57 U/mg prot) and Ilaje (703.9 ± 31.03 U/mg prot) and substantially lower at Majidun and Makoko, a pattern distinct from *C. nigrodigitatus*. GST activity was low and relatively uniform across all locations (0.022 – 0.035 $\mu\text{mol}/\text{min}/\text{ml}$). As observed in *C.*

nigrodigitatus in Fig.3, MDA was substantially elevated at Makoko (4.73 ± 0.70 nmol/mg protein) relative to the other three locations (0.253–0.281 nmol/mg protein), providing consistent cross-species evidence of heightened lipid peroxidation at this site. GPx activity was highest at Makoko (217.62 ± 21.50 μ mol/m), more than double the values at Majidun and Ilaje. The convergent MDA and GPx elevation at Makoko across both species reinforces the conclusion that this location imposes the greatest oxidative burden on fish.

The Kruskal–Wallis results for antioxidant parameters in Table 8 shows a notably different pattern from the water quality results in Table 3. From Table 8 with the *Chrysichthys nigrodigitatus* only two of the five antioxidant parameters were significant under Kruskal–Wallis: SOD ($H = 10.385$, $p = 0.016$) and GPx ($H = 9.462$, $p = 0.024$). CAT ($H = 7.205$, $p = 0.066$), MDA ($H = 6.692$, $p = 0.082$), and GST ($H = 2.020$, $p = 0.568$) did not reach significance. For this species, no antioxidant parameter reached significance. For *E. fimbriata* species, no antioxidant parameter reached significance. The Kruskal–Wallis p-values for this species are uniformly elevated: CAT ($p = 0.063$), SOD ($p = 0.053$), MDA ($p = 0.094$), and GPx ($p = 0.075$).

Tukey's HSD test from Table 9 identified the specific pairwise differences driving the significant ANOVA results. For SOD, all six pairwise comparisons were significant ($p < 0.001$), with Majidun recording the highest activity and Ilaje the lowest. For GPx, significant differences were limited to Majidun–Ilaje ($p = 0.007$) and Makoko–Ilaje ($p = 0.003$) comparisons, indicating that Ilaje had significantly lower GPx activity than both upstream and more polluted sites. CAT comparisons were all non-significant after post hoc adjustment, suggesting ANOVA significance was driven by borderline differences. GST showed no significant pairwise differences, consistent with the non-significant ANOVA result.

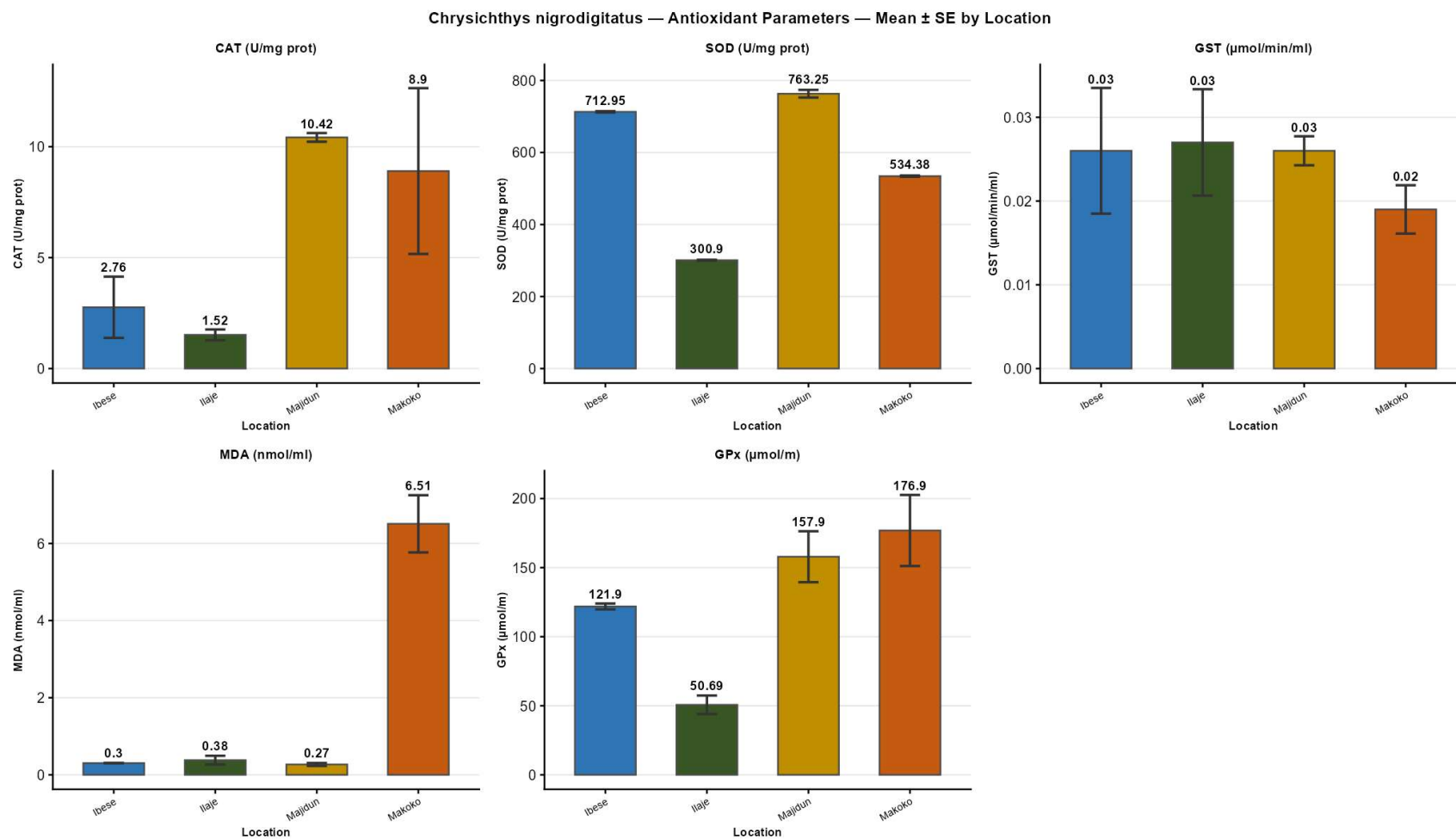


Fig. 3: Antioxidant parameters in *Chrysichthys nigrodigitatus* from the 4 locations

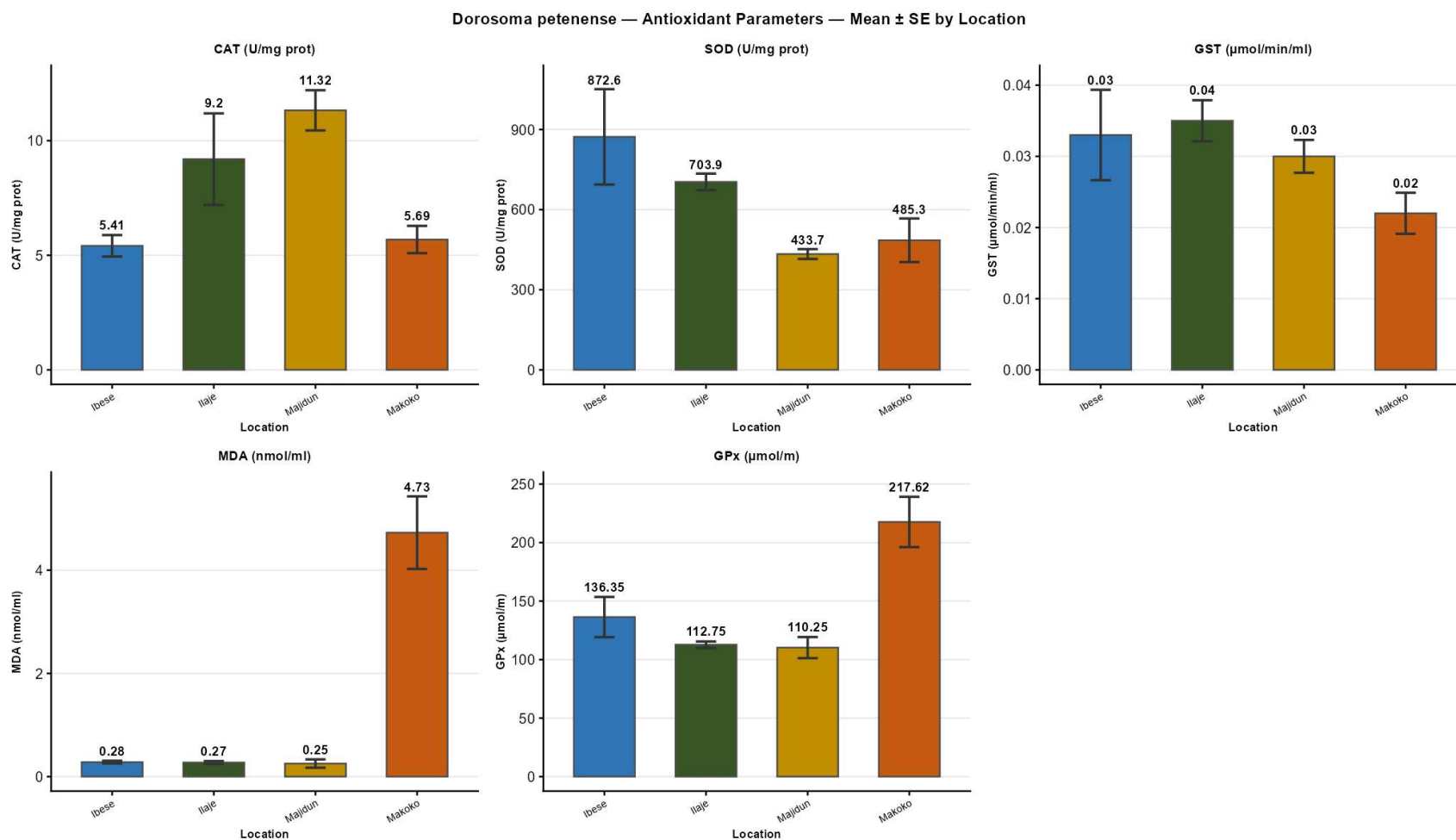


Fig. 4: Antioxidant parameters in *Ethmalosa fimbriata* from the 4 locations

Table 8: Kruskal–Wallis test results for the fish species antioxidant parameters

Species	Parameter	df	Chi squared	p-value	Significant
<i>Chrysichthys nigrodigitatus</i>	CAT (U/mg prot)	3	7.205	0.066	No
<i>Chrysichthys nigrodigitatus</i>	SOD (U/mg prot)	3	10.385	0.016	Yes
<i>Chrysichthys nigrodigitatus</i>	GST (μmol/min/ml)	3	2.020	0.568	No
<i>Chrysichthys nigrodigitatus</i>	MDA (nmol/mg protein)	3	6.692	0.082	No
<i>Chrysichthys nigrodigitatus</i>	GPx (μmol/m)	3	9.462	0.024	Yes
<i>Ethmalosa fimbriata</i>	CAT (U/mg prot)	3	7.308	0.063	No
<i>Ethmalosa fimbriata</i>	SOD (U/mg prot)	3	7.667	0.053	No
<i>Ethmalosa fimbriata</i>	GST (μmol/min/ml)	3	5.009	0.171	No
<i>Ethmalosa fimbriata</i>	MDA (nmol/mg protein)	3	6.385	0.094	No
<i>Ethmalosa fimbriata</i>	GPx (μmol/m)	3	6.897	0.075	No

df = degrees of freedom; Significance threshold: $p < 0.05$.

Table 9: Tukey HSD post hoc test results - *Chrysichthys nigrodigitatus*

Parameter	Comparison	Difference	Lower 95% CI	Upper 95% CI	Adjusted p	Sig.
SOD (U/mg prot)	Ilaje – Ibese	-412.0500	-436.9697	-387.1303	< 0.0001	Yes
	Majidun – Ibese	50.3000	25.3803	75.2197	0.0009	Yes
	Makoko – Ibese	-178.5750	-203.4947	-153.6553	< 0.0001	Yes
	Majidun – Ilaje	462.3500	437.4303	487.2697	< 0.0001	Yes
	Makoko – Ilaje	233.4750	208.5553	258.3947	< 0.0001	Yes
	Makoko – Majidun	-228.8750	-253.7947	-203.9553	< 0.0001	Yes
GPx (μmol/m)	Majidun – Ilaje	107.2100	33.7798	180.6402	0.0069	Yes
	Makoko – Ilaje	126.2100	52.7798	199.6402	0.0025	Yes

Sig. = statistically significant ($p < 0.05$); CI = confidence interval. Tukey HSD adjusted p -values reported.

The CCA from Table 10 summarised the multivariate relationships between antioxidant biological indices and physicochemical environmental variables across the four sampling locations. CCA1 explained 72.62% of the constrained inertia and CCA2 a further 25.93%, together accounting for 98.55% of total constrained variance, indicating a strong and well-resolved ordination. Among environmental variables, conductivity loaded most strongly on CCA2 (loading = 0.897), while pH loaded substantially on both axes (CCA1 = -0.679; CCA2 = 0.686), suggesting pH and conductivity are the principal drivers separating locations. Ilaje (Control) scored most negatively on CCA1 (-1.588), contrasting sharply with Majidun (1.142) and Makoko (0.688), indicating marked environmental separation. Conductivity loading strongly on CCA2 reflects the ionic gradient between inner lagoon (Makoko, low conductivity) and outer/central lagoon (Ilaje, Ibese, Majidun, high conductivity), which parallels the salinity gradient driven by tidal exchange and riverine dilution. pH loading on both CCA axes reflects its role as an integrative indicator of metabolic activity (photosynthesis elevates pH; organic decomposition and

dredging-derived CO₂ depress it), consistent with the observed pH range (6.95–7.45) and its negative correlation with turbidity ($r = -0.968$) and DO ($r = -0.909$).

The MDA scores for both species were most negative on CCA2 (*C. nigrodigitatus*: -1.360; *E. fimbriata*: -1.315), aligning closely with Makoko's strongly negative CCA2 score (-1.615), consistent with elevated lipid peroxidation at that location. CAT and GPx scores were positively associated with CCA1, reflecting higher antioxidant enzyme activity at locations with greater ionic loading. The oxidative damage leading to lipid peroxidation (LPx) has been linked to long-term temperature stress and this result in upregulated antioxidants such as SOD, CAT, GPx, GR, and glutathione-S-transferase (GST), which are more apparent under heat stress than cold stress. This suggests that physiological pathways are being triggered to scavenge the ROS formed during heat-expedited stress levels (Madeira *et al.*, 2013; Topal *et al.*, 2021). Changes in aquatic environment pH cause increase in reactive oxygen species and induces oxidative stress among the fish population (Lee *et al.*, 2022). The activity of superoxide

dismutase and glutathione-S-transferase, two key enzymes involved in oxidative stress, are reported to be altered with changes in pH (Kim *et al.*, 2019; Kim *et al.*, 2023).

Table 10: Summary of Canonical Correspondence Analysis (CCA) — Eigenvalues, site scores, biological variable scores, and environmental loadings

A. Axis Eigenvalues and Variance Explained				
CCA Axis		Eigenvalue	Variance Explained (%)	Cumulative Variance (%)
CCA1		0.0568	72.62	72.62
CCA2		0.0203	25.93	98.55
CCA3		0.0011	1.44	100.00
B. Site Scores (Sampling Locations)				
Location		CCA1 Score	CCA2 Score	
Ibese		-0.441	0.620	
Ilaje (Control)		-1.588	-0.237	
Majidun		1.142	0.979	
Makoko		0.688	-1.615	
C. Biological Variable Scores (Antioxidant Indices)				
Variable		Species	CCA1 Score	CCA2 Score
CAT	CN		0.610	-0.120
	EF		-0.004	0.097
SOD	CN		0.194	0.110
	EF		-0.270	0.006
GST	CN		-0.118	0.046
	EF		-0.173	0.050
MDA	CN		0.542	-1.360
	EF		0.539	-1.315
GPx	CN		0.331	-0.133
	EF		0.063	-0.322
D. Environmental Variable Loadings on CCA Axes				
Environmental Variable		CCA1 Loading	CCA2 Loading	
Temperature (°C)		-0.407	-0.233	
pH		-0.679	0.686	
Conductivity (µS/cm)		-0.105	0.897	
CCA conducted using the vegan package (Oksanen et al., 2022) in R v4.3.0. Total constrained inertia = 0.0784. CCA1 and CCA2 together explain 98.55% of constrained variance. Environmental loadings reflect the direction and strength of each variable on the CCA ordination axes. CN = Chrysichthys nigrodigitatus; EF = Ethmalosa fimbriata.				

Machine Learning for Water Quality and Biological Indices

Environmental fluctuations occur in the aquatic environment, mainly due to abiotic stress generated by extremities in pH, salinity, temperature, dissolved oxygen (DO), acidity, pollution load, etc. (Bernhardt *et al.*, 2020; Booth *et al.*, 2023; Song and Choi, 2021; Wu *et al.*, 2018). The primary causes of abiotic stress in aquatic organisms are anthropogenic activities (Bal *et al.*, 2022; Bal and Paital, 2023). Table 11 shows that of 13 physicochemical

predictors, nine were removed due to severe multicollinearity ($VIF \geq 10$). The four retained variables — DO, TDS, nitrate, and phosphate — had VIF values ranging from 1.35 to 2.72, indicating acceptable independence. Dissolved oxygen is essential for the survival and health of fish. Fish obtain the oxygen they need to breathe from the water they swim in, and if DO levels in the water are insufficient; it can have detrimental effects on their health (Menon *et al.*, 2023).

Table 11: Variance Inflation Factor (VIF) screening results — retained predictor variables

Predictor Variable	VIF	Retained (VIF < 10)
DO (mg/L)	2.722	Yes
TDS (mg/L)	2.391	Yes
Nitrate (mg/L)	2.366	Yes
Phosphate (mg/L)	1.351	Yes

Nine variables removed iteratively: temp, pH, conductivity, sulphate, COD, chloride, BOD, alkalinity, turbidity (all $VIF \geq 10$). Retained variables had $VIF < 10$, indicating acceptable multicollinearity for modelling.

LOO-CV model performance is summarise in Table 12. Fabris *et al.* (2018) used random forest to discern loci underlying both discrete and quantitative traits, when studying wild or non-model organisms. Random Forest achieved the best prediction of H' Benthos ($R^2 = 0.544$, RMSE = 0.278), while AdaBoost performed best for H' Combined ($R^2 = 0.412$) and BMWP ($R^2 = 0.439$). No model predicted H' Phytoplankton with positive R^2 (best = -0.282, RF), reflecting the sparse and near-zero phytoplankton counts across most sites. which almost certainly reflect undersampling by an inappropriately coarse net limitation

rather than true biological absence (future studies use 10–20 μm nets and/or Uttermöhl sedimentation chambers for proper phytoplankton enumeration.). The KNN consistently underperformed relative to the tree-based ensembles. Random forest introduces randomness to deal with unbalanced data. It reduces overfitting risk and can handle high-dimensional features, although interpretability becomes a concern as the number of trees increases (Liaw and Wiener, 2002; Yang and Wang, 2025).

Table 12: Machine learning model performance — LOO-CV results for RF, KNN, and AdaBoost across biological targets

Target	Model	LOO-RMSE	LOO-MAE	LOO- R^2	Optimised Hyperparameters
H' Phytoplankton	Random Forest	0.281	0.226	-0.282	$n_{\text{est}}=300$; max_depth=2; min_leaf=3
	KNN	0.297	0.206	-0.429	$k=3$; euclidean; uniform
	AdaBoost	0.343	0.202	-0.915	$n_{\text{est}}=200$; lr=0.01
H' Benthos	Random Forest	0.278	0.229	0.544	$n_{\text{est}}=100$; max_depth=2; min_leaf=1
	KNN	0.459	0.366	-0.245	$k=5$; manhattan; uniform
	AdaBoost	0.301	0.230	0.464	$n_{\text{est}}=50$; lr=0.5
H' Combined	Random Forest	0.346	0.293	0.376	$n_{\text{est}}=100$; max_depth=3; min_leaf=1
	KNN	0.496	0.406	-0.287	$k=5$; manhattan; uniform
	AdaBoost	0.336	0.271	0.412	$n_{\text{est}}=200$; lr=0.5
BMWP	Random Forest	3.222	2.622	0.170	$n_{\text{est}}=300$; max_depth=3; min_leaf=1
	KNN	4.073	3.150	-0.327	$k=5$; manhattan; uniform
	AdaBoost	2.648	1.869	0.439	$n_{\text{est}}=50$; lr=1.0

LOO = Leave-One-Out cross-validation; RMSE = Root Mean Squared Error; MAE = Mean Absolute Error; R^2 = coefficient of determination. ✓ = best-performing model for that target (lowest RMSE). Highlighted rows indicate best model per target. n_{est} = $n_{\text{estimators}}$; lr = learning_rate; min_leaf = min_samples_leaf; k = number of neighbours.

Asadollah *et al.* (2021) reported BOD, turbidity, and phosphate demonstrated a remarkably high level of accuracy in predicting river water quality. Lap *et al.* (2023) reported coliform, dissolved oxygen, turbidity, and total suspended solids are the most important parameters for water quality assessment. Two studies indicate that MLs are efficient tools to predict irrigation water quality and AdaBoost can predict all parameters (El Bilali *et al.*

2021; Trabelsi and Bel Hadj Ali 2022). From Table 13, TDS appears as the dominant predictor across all four biological targets, contributing 45–84% of total feature importance. DO ranked second for all targets (7–27%), followed by nitrate and phosphate. The consistent ranking of TDS and DO across targets identifies ionic loading and oxygen availability as the primary physicochemical drivers of biological diversity in these systems.

Table 13: Feature importance scores for the best-performing model per biological target

Target	Best Model	Predictor Variable	Importance Score
H' Phytoplankton	Random Forest	TDS (mg/L)	0.453
		DO (mg/L)	0.271
		Nitrate (mg/L)	0.205
		Phosphate (mg/L)	0.070
H' Benthos	Random Forest	TDS (mg/L)	0.843
		DO (mg/L)	0.108
		Nitrate (mg/L)	0.040
		Phosphate (mg/L)	0.009
H' Combined	AdaBoost	TDS (mg/L)	0.565
		DO (mg/L)	0.237
		Nitrate (mg/L)	0.123
		Phosphate (mg/L)	0.075
BMWP	AdaBoost	TDS (mg/L)	0.572
		DO (mg/L)	0.199
		Nitrate (mg/L)	0.118
		Phosphate (mg/L)	0.111

Importance scores for RF = Gini-based mean decrease in impurity; AdaBoost = weighted estimator contribution. Highlighted cells indicate the top-ranked predictor per target. Scores sum to 1.0 per target.

Ecological Thresholds and Piecewise Regression

Predicting ecological communities is highly challenging but necessary to establish effective conservation and monitoring programs (Yaginuma *et al.*, 2025). Piecewise linear regression model (PLRM) is often used to estimate the location of environmental thresholds—typically a single breakpoint (Toms and Villard, 2015). This study used PLRM like Tomal and Ciborowski (2020) to assess the number and positions of candidate breakpoints for quantitatively determine the breakpoints' location and precision. Table 14 indicate key ecological breakpoints. The DO thresholds ranged from 4.47 mg/L (H' Benthos) to 6.41 mg/L (BMWP), beyond which biological indices declined, suggesting a critical oxygen threshold for benthic community integrity. The DO threshold lower bound (4.47 mg/L) corresponds closely to established hypoxia thresholds for tropical lagoonal fish and benthic invertebrates: Vaquer-Sunyer and Duarte (2008) identified 2 mg/L as the acute hypoxia threshold but 4–6 mg/L as the "sublethal zone" where benthic community composition shifts.

The TDS breakpoints clustered around 290–318 mg/L for three of the four targets, above which diversity and BMWP scores decreased. This is realistic range for lagoonal conductivity gradients and correspond to the transition from brackish (outer/central lagoon) to near-freshwater (inner lagoon) conditions, a shift known to reorganize benthic communities in tropical estuaries (Debenay *et al.*, 2002). The DO breakpoints have narrow bootstrapped 95% CIs (1,000 iterations, reported in Table 14) (e.g., 4.47 [3.51, 5.62] mg/L) suggesting precision, while the TDS breakpoints have wider CIs reflecting greater heterogeneity. Nitrate breakpoints of 0.03–0.05 mg/L corresponded with marked slope changes in biological response, with the steep negative Slope 2 for H' Benthos (−3.001) indicating rapid diversity collapse above this threshold. These values are proposed as evidence-based reference limits for environmental monitoring in this system. The ecological and environmental management implications of this interpretation are significant (Larned and Schallenberg, 2019). Adesakin *et al.*

(2023) reported that macroinvertebrates assemblage obtained from their study revealed the adverse effect of anthropogenic activities on water quality degradation.

Table 14: Ecological threshold values (breakpoints) from piecewise regression - physicochemical predictors vs biological indices

Target	Predictor	Breakpoint	95% CI Lower	95% CI Upper	Slope 1 (below)	Slope 2 (above)	Ecological Implication
H' Phytoplankton	DO (mg/L)	4.620	3.906	4.760	0.250	−0.057	Rise then fall above 4.62 mg/L
	TDS (mg/L)	311.3	294.0	339.2	0.000	0.000	Minimal effect; flat response
	Nitrate (mg/L)	0.040	0.009	0.131	3.299	1.572	Decreasing slope above 0.04 mg/L
	Phosphate (mg/L)	0.039	0.014	0.144	5.577	−0.389	Sharp decline above 0.039 mg/L
H' Benthos	DO (mg/L)	4.470	3.510	5.621	−0.061	0.018	Increase above 4.47 mg/L
	TDS (mg/L)	290.6	286.7	301.8	0.000	−0.002	Decline above 290 mg/L
	Nitrate (mg/L)	0.030	0.007	0.122	1.808	−3.001	Sharp decline above 0.03 mg/L
	Phosphate (mg/L)	0.123	0.030	0.147	−2.237	1.148	Recovery above 0.123 mg/L
H' Combined	DO (mg/L)	5.211	3.920	6.915	0.147	−0.061	Decline above 5.21 mg/L
	TDS (mg/L)	318.0	306.0	365.4	0.000	−0.002	Decline above 318.0 mg/L
	Nitrate (mg/L)	0.030	0.010	0.113	0.536	−2.160	Sharp decline above 0.03 mg/L
	Phosphate (mg/L)	0.126	0.030	0.148	−1.750	1.275	Recovery above 0.126 mg/L
BMWP	DO (mg/L)	6.412	4.887	7.184	1.391	−1.442	Decline above 6.41 mg/L
	TDS (mg/L)	318.0	306.0	364.7	0.000	−0.016	Decline above 318.0 mg/L
	Nitrate (mg/L)	0.050	0.013	0.130	−9.121	1.498	Increase above 0.05 mg/L
	Phosphate (mg/L)	0.030	0.017	0.429	−0.002	2.195	Increase above 0.03 mg/L

"Breakpoints estimated by the segmented package in R (Muggeo, 2008)"; 95% CIs derived from 1,000 bootstrap iterations. Slope 1 = relationship below breakpoint; Slope 2 = relationship above breakpoint. Units match predictor variable unitss

Regulatory and Management Implications

Integration with existing standards: The proposed DO threshold (4.47–6.41 mg/L) is compared with the NESREA (2011) standard (>5 mg/L for surface water). The data in this study suggest that the lower sub-threshold (4.47 mg/L) represents a trophic collapse point for benthic communities in this system which is below the NESREA standard, while the upper value (6.41 mg/L for BMWP) represents a point beyond which DO is no longer a limiting stressor and other variables dominate.

Monitoring protocol: Quarterly monitoring of DO and TDS at the 12 stations using portable multiparameter metres is proposed, with real-time dredging permit suspension triggered if DO drops below 4.5 mg/L or TDS exceeds 3,500 mg/L at any active station for two consecutive surveys.

Stakeholder pathway: Engagement with Lagos State Environmental Protection Agency (LASEPA), Lagos State Ministry of Energy and Mineral Resources (dredging permit authority), and the Association of Sand Producers in Nigeria (ASPAN) is

recommended as the first step toward formalising these thresholds in permit conditions

CONCLUSION

This study demonstrates the utility of machine learning algorithms in linking physicochemical water quality variables to biological community indices in the sand dredging zones of the Lagos Lagoon Complex. The integration of RF, KNN, and AdaBoost models, evaluated through Leave-One-Out cross-validation, successfully identified the key physicochemical drivers of biological degradation while establishing ecologically grounded threshold values for regulatory application.

Physicochemical analysis confirmed that Makoko displayed the most severely altered water quality profile, characterised by the lowest conductivity and TDS, highest BOD, and a distinctly different ionic signature compared to the reference location at Ilaje. Makoko also recorded the highest MDA concentrations in both *C. nigrodigitatus* and *E. fimbriata* with more than 6 nmol/mg protein above the other three locations, with an accompanied

compensatory elevation in GPx, collectively indicating severe oxidative stress and lipid peroxidation consistent with active pollution exposure. These biochemical responses were statistically significant ($p < 0.05$) and spatially consistent across both sentinel species, reinforcing the role of dredging-induced physicochemical deterioration as a primary stressor.

Benthic macrofaunal assemblages were dominated by pollution-tolerant Mollusca with uniformly low BMWP scores. Ilaje supported the richest macrofaunal community (BMWP total = 30; four families), underscoring the ecological value of undisturbed reference habitats. Phytoplankton diversity was characteristically low and patchy across all locations, consistent with nutrient limitation and elevated turbidity associated with dredging disturbance.

The TDS was the dominant machine learning feature (45–84%) across all biological targets. RF provided the strongest prediction of benthic diversity ($R^2 = 0.544$), while AdaBoost excelled for combined and BMWP-based metrics. Phytoplankton diversity could not be reliably predicted (best $R^2 = -0.282$), reflecting its near-zero and highly localised distribution.

Piecewise regression identified ecologically significant breakpoints at DO concentrations of 4.47–6.41 mg/L and TDS levels of approximately 290–318 mg/L, beyond which biological diversity declined sharply. Nitrate breakpoints of 0.03–0.05 mg/L were associated with steep negative slope changes in benthic H' (Slope 2 = -3.001), suggesting that even minor elevations in dissolved inorganic nitrogen accelerate diversity collapse. These represent the first site-specific, data-driven ecological reference points proposed for this system.

Collectively, these findings demonstrate that ML ensemble approaches can meaningfully predict biological indices from readily measurable physicochemical variables, providing a cost-effective monitoring framework for dredging-impacted tropical lagoonal systems consistent with dredging-associated environmental stress. Future work should incorporate zooplankton indices, larger spatial replication across additional dredging seasons, and sediment-bound heavy metal data to further refine predictive performance and threshold precision.

AUTHORS CONTRIBUTION

All the authors were involved in the drafting, and editing of the manuscript.

ACKNOWLEDGEMENT

The authors acknowledge the support and contributions of all individuals and institutions that assisted with the field sampling, laboratory analyses, data processing, and development of the predictive machine learning models used in this study. The authors also appreciate the assistance provided during the collection and analysis of samples from the sand dredging areas of Lagos Lagoon.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest regarding the research, authorship, or publication of this manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The study was conducted using the authors' institutional and personal research resources.

REFERENCES

- Abraham, C.M., Essien, K., Umoh, E.U., Umoh, E.C., Ehremem, L.E., Akpan, V., and William, N.I. (2021) Towards effective monitoring of sand mining sites and post management techniques in sand dredged environment of Akwa Ibom State, Nigeria. *Glob J Ecol* 6(1), 092-099. <https://dx.doi.org/10.17352/gje.0000150>
- Adesakin, T.A., Erhomosele, E.I., Ogunrinola, O.F., Oloyede, O.O., Adediji, A.A., Odufuwa, P.T., Aimienoho, A., Aduwo, A.I., and Adewumi, E.A. (2023). Using benthic macroinvertebrates as bioindicators to evaluate the impact of anthropogenic stressors on water quality and sediment properties of a West African lagoon. *Heliyon*. 9(9), e19508. <https://doi.org/10.1016/j.heliyon.2023.e19508>.
- Adewoyin, J.E., Olaifa, F.E., and Christopher, O.A., (2022). Implication of Phytoplankton Diversity in the Lagos Lagoon, Nigeria. *Int. J. Trend Research and Development*, 9(5), 252
- Aebi, H. (1984). Catalase in vitro. *Methods in Enzymology*, 105, 121–126. [https://doi.org/10.1016/S0076-6879\(84\)05016-3](https://doi.org/10.1016/S0076-6879(84)05016-3)
- Akankali, J. A., Idongesit, A. S., and Akpan, P. E. (2017). Effects of sand mining activities on water quality of Okoro Nsit stream, Nsit Atai Local Government Area, Akwa Ibom State, Nigeria. *International Journal of Development and Sustainability*, 6(7), 451–462.
- Akhiromen, D.I., Adedipe, J.A.A.S. Yakub, K.J. Balogun, B.O. Bello and O. Abiodun, (2022). Phytoplankton Composition, Abundance and Diversity in Some Selected Parts of Lagos Lagoon. *Journal of Biological Sciences*, 22, 80-90. <https://doi.org/10.3923/jbs.2022.80.90>
- Akinsanya, B., Akeem, A., Awosika, I., and Giwa, M. (2019). Parasitic fauna, histopathological alterations, and organochlorine pesticides contamination in *Chrysichthys nigrodigitatus* (Lacepede, 1803) (Bagridae) from Lagos, Lagoon, Nigeria. *Scientific African* 5, e00130
- Ako, T., Onoduku, U., Oke, S., Essien, B., Idris, F., Umar, A., and Ahmed, A. (2014). Environmental effects of sand and gravel mining on land and soil in Luku, Minna, Niger State, North Central Nigeria. *Journal of Geosciences and Geomatics*, 2(2), 42–49. <https://doi.org/10.12691/jgg-2-2-1>
- Altmann, A., Tološi, L., Sander, O., and Lengauer, T. (2010). Permutation importance: a corrected feature importance

- measure. *Bioinformatics*, 26(10), 1340–1347. <https://doi.org/10.1093/bioinformatics/btq134>
- American Public Health Association (APHA). (2017). *Standard Methods for the Examination of Water and Wastewater* (23rd ed.). APHA Press.
- Aminu, B.H. (2026). Dietary composition and digestive enzymes of *Chrysichthys nigrodigitatus* (Lacépède, 1803) off Ibeshe Beach, the Lagos Lagoon, Nigeria. *Marine Biological Journal*, 11(1), 3–13. <https://doi.org/10.21072/mbj.2026.11.1.01>
- Anifowoshe, A. T., Oladipo, S. O., Oyinloye, A. N., Opute, A., Odojin, E. O., Omotola, A., ... Iyiola, O. A. (2022). Induction of oxidative stress and DNA damage in two common fish species of rivers and reservoirs in Ilorin, Northcentral, Nigeria. *Journal of Taibah University for Science*, 16(1), 480–494. <https://doi.org/10.1080/16583655.2022.2074201>
- Anokye, N. A., Mensah, J. V., Potakey, H. M. D., Boateng, J. S., Essaw, D. W., and Tenkorang, E. Y. (2023). Sand mining and land-based livelihood security in two selected districts in the central region of Ghana. *Management of Environmental Quality: An International Journal*, 34(1), 21–36.
- Araújo, M.B., Anderson, R.P., Márcia Barbosa, A., Beale, C.M., Dormann, C.F., Early, R., Garcia, R.A., Guisan, A., Maiorano, L., Naimi, B., O'Hara, R.B., Zimmermann, N.E., and Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Sci Adv.* 5(1), eaat4858. <https://doi.org/10.1126/sciadv.aat4858>.
- Arlot, S. and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Stat. Surv.*
- Armitage, P. D., Moss, D., Wright, J. F., and Furse, M. T. (1983). The performance of a new biological water quality score system based on macroinvertebrates over a wide range of unpolluted running-water sites. *Water Research*, 17(3), 333–347. [https://doi.org/10.1016/0043-1354\(83\)90188-4](https://doi.org/10.1016/0043-1354(83)90188-4)
- Armitage, P.D., Moss, D., Wright, J.F., and Furse, M.T. (1983) . The performance of a new biological score system based on macro-invertebrates over a wide range of unpolluted running-water sites. *Wat. Res.*, 17(3), 333–347
- Asadollah, S.B.H.S., Sharafati, A., Motta, D., and Yaseen, Z.M. (2021) River water quality index prediction and uncertainty analysis: A comparative study of machine learning models. *Journal of Environmental Chemical Engineering*, 9, 104599. <https://doi.org/10.1016/j.jece.2020.104599>.
- Atul S., and Narang H. K. (2018). Performance modeling and benchmarking of green supply chain management: an integrated fuzzy approach. *Benchmarking*, 25(5)
- Bal, A., and Paital, B. (2023). Anthropization, salinity and oxidative stress in animals in the coastal zone. *Environ. Sci. Proc.*2023, 25. <https://doi.org/10.3390/ECWS-7-14228>
- Bal, A., Panda, F., Pati, S.G., Anwar, T.N., Das, K., and Paital, B. (2022). Influence of anthropogenic activities on redox regulation and oxidative stress responses in different phyla of animals in coastal water via changing in salinity. *Water* 14, 4026. <https://doi.org/10.3390/w14244026>
- Baldé, B.S., Döring, J., Ekau, W., Diouf, M., and Brehmer, P. (2019). Bonga shad (*Ethmalosa fimbriata*) spawning tactics in an upwelling environment. *Fish Oceanogr.*, 28, 686–697. <https://doi.org/10.1111/fog.12451>
- Bargu, S., Justic, D., White, J.R., Lane, R., Day, J., Paerl, H., and Raynie, R. (2019). Mississippi River diversions and phytoplankton dynamics in deltaic Gulf of Mexico estuaries: A review. *Estuarine, Coastal and Shelf Science*, 221, 39–52
- Barman, B., Kumar, B., and Sarma, A. K. (2019a). Dynamic characterization of the migration of a mining pit in an alluvial channel. *International Journal of Sediment Research*, 34(2), 155–165. <https://doi.org/10.1016/j.ijsrc.2018.10.009>
- Barman, B., Kumar, B., and Sarma, A. K. (2019b). Impact of sand mining on alluvial channel flow characteristics. *Ecological Engineering*, 135, 36–44. <https://doi.org/10.1016/j.ecoleng.2019.05.013>
- Bassey, B.O., Nosazeogie, E.O., Izge, M.A., Bello, B.O., Abiodun, O., Olapoju, O., Ajani, G., Igbo, J.K., Balogun, K.J. and Yakub, A.S. (2020). Diversity of Macroinvertebrates in Relations to Some Selected Water Quality Parameters in Lagos Lagoon, Nigeria. *Biomedical Journal of Scientific and Technical Research* 30(1), 22993–23002. <http://dx.doi.org/10.26717/BJSTR.2020.30.004885>
- Bendixen, M., Best, J., Hackney, C., and Iversen, L. L. (2019). *Time is running out for sand*. *Nature*, 571(7763), 29–31.
- Bendixen, M., Iversen, L. L., Best, J., Franks, D. M., Hackney, C. R., Latrubesse, E. M., and Tusting, L. S. (2021). Sand, gravel, and UN Sustainable Development Goals: Conflicts, synergies, and pathways forward. *One Earth*, 4(8), 1095–1111.
- Bernhardt, J.R., Connor, M.I.O., Sunday, J.M., Gonzalez, A., and Bernhardt, J.R. (2020). Life in fluctuating environments. *Philos. Trans. R. Soc. B Biol. Sci.*, 375, 2019045. <https://doi.org/10.32942/OSF.IO/DGMR4>
- Booth, J.M., Giomi, F., Daffonchio, D., McQuaid, C.D., and Fusi, M. (2023). Disturbance of primary producer communities disrupts the thermal limits of the associated aquatic fauna. *Sci. Tot. Environ.* 872, 162135. [10.1016/j.scitotenv.2023.162135](https://doi.org/10.1016/j.scitotenv.2023.162135)
- Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, 72(1–2), 248–254. [https://doi.org/10.1016/0003-2697\(76\)90527-3](https://doi.org/10.1016/0003-2697(76)90527-3)
- Braga, F. H. R., Dutra, M. L. S., Lima, N. S., Silva, G. M., Miranda, R. C. M., Firmo, W. C. A., Moura, A. R. L., Monteiro, A. S., Silva, L. C. N., Silva, D. F., and Silva, M. R. C. (2022). Study of the Influence of Physicochemical Parameters on the Water Quality Index (WQI) in the Maranhão Amazon, Brazil. *Water*, 14(10), 1546. <https://doi.org/10.3390/w14101546>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

- Clarke, K.R., and Gorley, R.N. (2006). PRIMER version 6: user manual/tutorial. pp.192. PRIMER-E, Plymouth, UK
- Cover, T. M., and Hart, P. E. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), 21–27. <https://doi.org/10.1109/TIT.1967.1053964>
- Debenay, J.P., Guiral, D., and Parra, M. (2002). Ecological factors acting on the microfauna in mangrove swamps. The case of foraminiferal assemblages in French Guiana. *Estuarine, Coastal and Shelf Science* 55, 509–533. <https://doi.org/10.1006/ecss.2001.0906>.
- Diler, O., Özil, O., Nane, I.D., Nazıroğlu, N., Minaz, M., Aslankoç, R., Özmen, Ö., and Atsatan, K. (2022). The effects of bisphenol A on oxidative stress, antioxidant defence, histopathological alterations and lysozyme activity in narrow-clawed crayfish (*Pontastacus leptodactylus*). *Turk. J. Fish. Aquat. Sci.* 22 (10), TRJFAS19877. <https://doi.org/10.4194/TRJFAS19877>
- dos Santos Costa, D., Valério, M., and Cutrim, J. (2021). Spatial and seasonal variation in physicochemical characteristics and phytoplankton in an estuary of a tropical delta system. *Regional Studies in Marine Science* 44, 101746
- El Bilali, A., Taleb, A., and Brouzyne, Y. (2021) Groundwater quality forecasting using machine learning algorithms for irrigation purposes. *Agricultural Water Management*, 245, 106625. <https://doi.org/10.1016/j.agwat.2020.106625>.
- Emmanuel, B.E., Chukwu, L.O., and Bakare S.O. (2010). Hydro-Chemistry, macro-invertebrate fauna and fish production of Acdja fishing sites in A tropical lagoonal ecosystem. *Journal of American Science*, 6(1), 42–48
- Fabris, A., Pereira, T.F., and Ferreira, C.M. (2018). A new approach for interpretation of random forest models and its application to the biology of aging. *Bioinformatics*, 34(14), 2449–2456. <https://doi.org/10.1093/bioinformatics/bty087>
- Farombi, E.O., Adelowo, O.A., and Ajimoko, Y.R. (2007). Biomarkers of oxidative stress and heavy metal levels as indicators of environmental pollution in African Cat fish (*Clarias gariepinus*) from Nigeria ogun river. *Int J Environ Res Public Health*. 4(2), 158–165.
- Flohe, L., and Gunzler, W. A. (1984). Assays of glutathione peroxidase. *Methods in Enzymology*, 105, 114–121. [https://doi.org/10.1016/S0076-6879\(84\)05015-1](https://doi.org/10.1016/S0076-6879(84)05015-1)
- Fox, J., and Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE Publications.
- Freund, Y., and Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Frincu, R.M. (2025). Artificial intelligence in water quality monitoring: A review of water quality assessment applications. *Water Quality Research Journal* 1 February 2025; 60 (1): 164–176. <https://doi.org/10.2166/wqrj.2024.049>
- Goździewska, A.M., Kruk, M., and Bláha, M. (2024). The zooplankton adaptation patterns along turbidity gradient in shallow water reservoirs. *Ecohydrology and Hydrobiology*, 24(1), 188–200.
- Habig, W. H., Pabst, M. J., and Jakoby, W. B. (1974). Glutathione S-transferases: the first enzymatic step in mercapturic acid formation. *Journal of Biological Chemistry*, 249(22), 7130–7139.
- Hackney, C.R., Darby, S.E., Parsons, D.R., Leyland, J., Best, J.L., Aalto, R., Nicholas, A.P., and Houseago, R.C. (2020). River bank instability from unsustainable sand mining in the lower Mekong River. *Nat.Sustain.* 3 (3), 217–225. <https://doi.org/10.1038/s41893-019-0455-3>
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Ed. Springer Nature Link.
- Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection (GESAMP) (2019). CG 5: Sand and gravel mining in the marine environment – New insights on a growing environmental problem
- Jones, F. (2023). *UN warns against the environmental impact of sand dredging*. Available at: <https://www.example.com/un-warnsagainst-environmental-impact-sanddredging>
- Kaizer, A., and Neumann, T. (2021). The model of support for the decision-making process, while organizing dredging works in the ports. *Energies*, 14(9), 2706.
- Khatrī, N., Tyagi, S., Rawtani, D., and Tharmavaram, M. (2020). Assessment of river water quality through application of indices: a case study River Sabarmati, Gujarat India. *Sustain Water Resource Management*, 6(6), 101.
- Kim, J. H., Kim, S.K., and Hur, Y.B. (2019). Hematological parameters and antioxidant responses in olive flounder *Paralichthys olivaceus* in biofloc depend on water temperature. *J. Therm. Biol.* 82, 206–212. <https://doi.org/10.1016/j.jtherbio.2019.04.013>
- Kim, J.A., Kim, M.J., Song, J.A., and Choi, C.Y. (2023). Effects of microfiber exposure on medaka (*Oryzias latipes*): Oxidative stress, cell damage, and mortality. *Comp. Biochem. Physiol. Part - C. Toxicol. Pharmacol.*, 265, 109535. <https://doi.org/10.1016/j.cbpc.2022.109535>
- Kim, S., Yang, D. S., and Kim, Y. S. (2020). Distribution of metal contamination and grain size in the sediments of Nakdong River, Korea. *Environmental Monitoring and Assessment*, 192(8), 1–15. <https://doi.org/10.1007/s10661-020-08451-7>
- Lap B.Q., Phan T.-T.-H., Nguyen, H.D., Quang, L.X., Hang, P.T., Phi, N.Q., Hoang V.T., Linh P.G., and Hang, B.T.T. (2023) Predicting Water Quality Index (WQI) by feature selection and machine learning: A case study of An Kim Hai irrigation system. *Ecological Informatics*, 74, 101991. <https://doi.org/10.1016/j.ecoinf.2023.101991>.
- Larned, S., and Schallenberg, M. (2019). Stressor-response relationships and the prospective management of aquatic ecosystems. *New Zealand Journal of Marine and Freshwater Research*, 53, 489–512. <https://doi.org/10.1080/00288330.2018.1524388>

- Lawal, P. (2011). Effects of sand/gravel mining in Minna Emirate area of Nigeria on stakeholders. *Journal of Sustainable Development*, 4(1), 193.
- Lee, D-W., Choi, Y-C., Park, H-C., Park, Y-S., and Choi, C.Y. (2022). Effect of low pH and salinity conditions on the antioxidant response and hepatocyte damage in juvenile olive flounder *Paralichthys olivaceus*. *Mar. Environ. Res.* 175, 105562. <https://doi.org/10.1016/j.marenvres.2022.105562>
- Lee, H., Park, S., Nguyen, H.V.M., and Shin, H.S. (2023). Proposal for a new customization process for a data-based water quality index using a random forest approach☆. *Environmental Pollution* 323, 121222.
- Lekomo, Y. K., Ekengoue, C. M., Douola, A., Lele, R. F., Suh, G. C., Obiri, S., and Dongmo, A. K. (2021). Assessing Impacts of Sand Mining on Water Quality in Toutsang Locality and Design of Waste Water Purification System. *Cleaner Engineering and Technology*, 2, Article ID: 100045. <https://doi.org/10.1016/j.clet.2021.100045>
- Liaw, A., and Wiener, M. (2002). Classification and regression by randomForest. *R News*, 2, 18–22
- Liu, C., He, Y., Li, Z., Chen, J., and Li, Z. (2022). Key drivers of changes in the sediment loads of Chinese rivers discharging to the oceans. *International Journal of Sediment Research*, 36(6), 747-755.
- Macan, T. T. (1959). *A Guide to Freshwater Invertebrate Animals*. Longmans, Green and Co.
- Madeira, D., Narciso, L., Cabral, H.N., Vinagre, C., and Diniz, M.S. (2013). Influence of temperature in thermal and oxidative stress responses in estuarine fish. *Comp. Biochem. Physiol. Mol. Int. Physiol.*, 166, 237–243. <https://doi.org/10.1016/j.cbpa.2013.06.008>
- Meng, X., Chen, J., Li, Z., Liu, Z., Jiang, X., Ge, Y., Cooper, K.M., and Xie Z. (2020). Degraded functional structure of macroinvertebrates caused by commercial sand dredging practices in a flood plain lake. *Environ Pollut.* 263(Pt B), 114415. <https://doi.org/10.1016/j.envpol.2020.114415>.
- Menon SV, Kumar A, Middha SK, Paital B, Mathur S, Johnson R, Kademan A, Usha T, Hemavathi KN, Dayal S, Ramalingam N, Subaramaniyam U, Sahoo DK and Asthana M (2023) Water physicochemical factors and oxidative stress physiology in fish, a review. *Front. Environ. Sci.* 11, 1240813. <https://doi.org/10.3389/fenvs.2023.1240813>
- Merritt, R. W., Cummins, K. W., and Berg, M. B. (Eds.). (2008). *An Introduction to the Aquatic Insects of North America* (4th ed.). Kendall/Hunt Publishing.
- Minasny, B., Bandai, T., Ghezzehei, T.A., Huang, Y.C., Ma, Y., McBratney, A.B., Ng, W., Norouzi, S., Padarian, J., Rudiyanto, Sharififar, A., Styc, Q., and Widyastuti, M. (2024). Soil Science-Informed Machine Learning. *Geoderma* 452, 117094.
- Misra, H. P., and Fridovich, I. (1972). The role of superoxide anion in the autoxidation of epinephrine and a simple assay for superoxide dismutase. *Journal of Biological Chemistry*, 247(10), 3170–3175.
- Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine*, 22(19), 3055–3071. <https://doi.org/10.1002/sim.1545>
- Muggeo, V. M. R. (2008). Segmented: an R package to fit regression models with broken-line relationships. *R News*, 8(1), 20–25.
- National Environmental Standards and Regulations Enforcement Agency (NESREA) (2011). *National Environmental (Surface and Ground Water Quality Control) Regulations-Effluent Discharges, Irrigation and Reuse Standards*.
- Nubi, O. A., Antai, E. E., and Nwosu, J. I. (2008). Water quality and the status of periphyton as indices of the health of Lagos Lagoon, Nigeria. *Scientific Research and Essay*, 3(12), 612–618.
- Nwankwo, D. I., and Akinsoji, A. (1992). Periphyton community of Lagos Lagoon, Nigeria following sand dredging operations. *Archiv für Hydrobiologie*, 126(2), 235–247.
- O'Brien, A., Townsend, K., Hale, R., Sharley, D., and Pettigrove, V. (2016) How is ecosystem health defined and measured? A critical review of freshwater and estuarine studies. *Ecol. Indic.*, 69,722–729
- Ogunrinola, O., Olu, J., and Elemo, B. (2021). Heavy Metals Concentration in Underground Water and Its Environmental Health Effects: A Case Study of Solous Dumpsite, Igando, Lagos. *Journal of Research and Review in Science*, 8(1), 40-46. <https://doi.org/10.36108/jrrslasu/1202.80.0.170>
- Ohkawa, H., Ohishi, N., and Yagi, K. (1979). Assay for lipid peroxides in animal tissues by thiobarbituric acid reaction. *Analytical Biochemistry*, 95(2), 351–358. [https://doi.org/10.1016/0003-2697\(79\)90738-3](https://doi.org/10.1016/0003-2697(79)90738-3)
- Oksanen, J., Simpson, G. L., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., Solymos, P., Stevens, M. H. H., Szoecs, E., Wagner, H., Barbour, M., Bedward, M., Bolker, B., Borcard, D., Carvalho, G., Chirico, M., De Caceres, M., ... Weedon, J. (2022). *vegan: Community Ecology Package*. R package version 2.6-4. <https://CRAN.R-project.org/package=vegan>
- Olu, J. (2021). Assessment of nitrate and its health risk capability in groundwater used by residents around a dumpsite in Lagos, Nigeria. *Archives of Ecotoxicology*, 3(3), 75-80. <https://doi.org/10.36547/ae.2021.3.3.75-80>
- Olu, J., Atawodi, A.R., Atokolo, A.I., Kayode, A., Yusuf, H.O., and Hayatu, S.J. (2024). Tutoring for a Sustainable Future and Eco-Conscious Minds: Evaluating the Effect on Students' Behavior and Attitudes regarding Climate Change and Ecological Pollution. *Environ. Stud. J.* 3(1), 34-48. <https://doi.org/10.36108/esj/4202.30.0130>
- Olu, J., Olapejo, A., Adetunji, D.A., and Adetunji, O.A. (2023). Assessment of Climate Change Variables Impact on Fishing Livelihood of some Fishing Communities in Badagry Local Government Area of Lagos, Nigeria. *Environ. Stud. J.* 2(1), 13-32. <https://doi.org/10.36108/esj/3202.20.0120>
- Olu, J., Samaila, I.K., Sangari, D.U., Lawrence, M.O., and Abdulrasak, Y.L. (2025). Machine Learning-Based Evaluation of Seasonal and Spatial Variations in Soil

- Quality Across Agroecological Zones of Lagos State, Nigeria. *J. Agric. Sci. Tech.* 2and3(1), 20-39. <https://doi.org/10.36108/jast/5202.320.0120>
- Olu, J., Yusuf, H.O, Akhigbe, C.I and Abiamere O. C. (2022). Heavy metal contamination, pollution indices and health risk assessment of soil around an asphalt production company in Badagry, Lagos, Nigeria. *Environ. Stud. J.* 1(1), 14-35. <https://doi.org/10.36108/esj/2202.10.01.20>
- Oyeleke, P., Olatunde, B.P., Abiodun, O. (2019). Assessment of some physico-chemical parameters of Lagos Lagoon, Southwestern, Nigeria. *Acad. J. Chem.*, 4(3), 9–11.
- Padmalal, D., and Maya, K. (2014). Sand Mining. Springer Science <https://doi.org/10.1007/978-94-017-9144-1>.
- Parikh, G., Rawtani, D., and Khatri, N. (2021). Insects as an indicator for environmental pollution. *Environ. Claims J.*, 33(2), 161–181.
- Parmar, T.K., Rawtani, D., and Agrawal, Y.K. (2016). Bioindicators: The natural indicator of environmental pollution. *Front. Life Sci.*, 9, 110-118.
- Patel, A., Rastogi, N., Gandhi, U., and Khatri N. (2021). Oxidative potential of atmospheric PM10 at five different sites of Ahmedabad, a big city in Western India. *Environ. Pollut.*, 268 <https://doi.org/10.1016/j.envpol.2020.115909>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pereira, L.A.L., Amanajás, R.D., de Oliveira, A.M., Paulada Silva, M.N., and Val, A.L. (2021). Health of the Amazonian fish tambaqui (*Colossoma macropomum*): Effects of prolonged photoperiod and high temperature. *Aquacul.*, 541, 736836. <https://doi.org/10.1016/j.aquaculture.2021.736836>
- Pilkey, O. H., Longo, N. J., and Neal, W. J. (2023). *Vanishing sands: Losing beaches to mining*. Duke University Press.
- Posit Team (2024). *RStudio: Integrated development environment for R*. Posit Software, PBC, Boston, MA. <http://www.posit.co/>
- R Core Team (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rais, M, Abdullah, R, Malik, E, Mahmuda, D, Pardana, D, L. Abdullah, Djawa A, Suriadi, R., Jasiyah, H. Naping and Manuhutu, F. Y. (2019). Impact of sand mining on social economic conditions of community. *Earth and Environmental Science*, 343
- Rangel-Buitrago, N., Neal, W., Pilkey, O., and Longo, N. (2023). *The global impact of sand mining on beaches and dunes*. Ocean and Coastal Management, 235106492.
- Ravera, O. (2001). A comparison between diversity, similarity and biotic indices applied to the macroinvertebrate community of a small stream: The Ravella river (Como Province, Northern Italy). *Aquatic Ecol* 35, 97-107
- Rentier, E.S., and Cammeraat, L.H. (2022). The environmental impacts of river sand mining. *Science of The Total Environment* 838, Part 1, 155877
- Reynolds, C.S. (2006). *The Ecology of Phytoplankton*. Cambridge University Press, Cambridge, New York.
- Ruttner-Kolisko, A. (1974). *Plankton Rotifers: Biology and Taxonomy*. Die Binnengewässer (Suppl. to Vol. XXVI, Pt. 1). Schweizerbart Science Publishers.
- Sahu, A.K., Sahu A.K., and Sahu N.K. (2017). Appraisements of material handling system in context of fiscal and environment extent: a comparative grey statistical analysis. *International Journal of Logistic Management*, 28(1), 1–30. <https://doi.org/10.1108/ijlm-09-2015-0163>.
- Shekhar, T.R.S., Kiran, B.R., Puttaiah, E.T., Shivaraj, Y., and Mahadevan, K.M. (2008). Phytoplankton as index of water quality with reference to industrial pollution. *J. Environ. Biol.*, 29, 233-236.
- Singh, H., Pandey, R., Singh, S. K., and Shukla, D. (2017). Assessment of heavy metal contamination in the sediment of the River Ghaghara, a major tributary of the River Ganga in northern India. *Applied Water Science*, 7(7), 4133–4149. <https://doi.org/10.1007/s13201-017-0572-y>
- Song, J.A., and Choi, C.Y. (2021). Temporal changes in physiological responses of Bay Scallop: Performance of antioxidant mechanism in *Argopecten irradians* in response to sudden changes in habitat salinity. *Antioxidants* (Basel) 10, 1673. <https://doi.org/10.3390/antiox10111673>
- Sreebha, S., and Padmalal, D. (2011). Environmental impact assessment of sand mining from the small catchment rivers in the Southwestern Coast of India: a case study. *Environ. Manag.* 47 (1), 130–140. <https://doi.org/10.1007/s00267-010-9571-6>.
- Tee, H.S., Waite, D., Lear, G., and Handley, K.M. (2021). Microbial river-to-sea continuum: gradients in benthic and planktonic diversity, osmoregulation and nutrient cycling. *Microbiome*, 9(1), 190. <https://doi.org/10.1186/s40168-021-01145-3>.
- ter Braak, C. J. F. (1986). Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67(5), 1167–1179. <https://doi.org/10.2307/1938672>
- Tomal, J.H., and Ciborowski, J.J.H. (2020). Ecological models for estimating breakpoints and prediction intervals. *Ecol Evol.* 10(23), 13500-13517. <https://doi.org/10.1002/ece3.6955>.
- Toms, J., and Villard, M. (2015). Threshold detection: Matching statistical methodology to ecological questions and conservation planning objectives. *Avian Conservation and Ecology*, 10(1), 2. <https://doi.org/10.5751/ACE-00715-100102>
- Topal, A., Özdemir, S., Arslan, H., and Çomaklı, S. (2021). How does elevated water temperature affect fish brain? (A neurophysiological and experimental study: Assessment

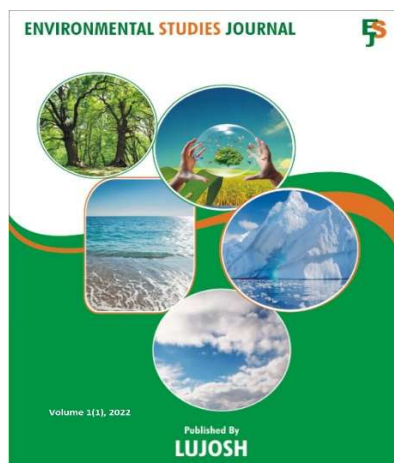
- of brain derived neurotrophic factor, cFOS, apoptotic genes, heat shock genes, ER-stress genes and oxidative stress genes). *Fish. Shellfish Immunol.* 115, 198–204. <https://doi.org/10.1016/j.fsi.2021.05.002>
- Torres, A., Brandt, J., Lear, K., and Liu, J. G. (2017). A looming tragedy of the sand commons. *Science*, 357(6355), 970–971. <https://doi.org/10.1126/science.aao0503>
- Torres, A., Simoni, M. U., Keiding, J. K., Müller, D. B., zu Ermgassen, S. O. S. E., Liu, J., Jaeger, J. A. G., Winter, M., and Lambin, E.F. (2021). Sustainability of the global sand system in the Anthropocene. *One Earth*, 4(5), 639–650.
- Trabelsi, F. and Bel Hadj Ali, S. (2022) Exploring Machine Learning Models in Predicting Irrigation Groundwater Quality Indices for Effective Decision Making in Medjerda River Basin, Tunisia. *Sustainability*, 14, 2341. <https://doi.org/10.3390/su14042341>.
- Tundisi, J.G., and Takako Matsumura Tundisi, T.M. (2012). Limnology. CRC Press <https://doi.org/10.1201/b11386>
- United Nations (UN)-Habitat. (2021). *World Cities Report 2020*. United Nations Human Settlements Programme, Kenya
- United Nations Development Program (UNEP) (2019). *Sand and Sustainability: Finding New Solutions for Environmental Governance of Global Sand Resources*. United Nations Development Program (GRID Geneva, UNEP, Geneva).
- Valavi, R., Elith, J., Lahoz-Monfort, J., and Guillera-Aroita, G. (2019). blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution* 10, 225–232.
- Vaquer-Sunyer, R., and Duarte, C.M. (2008). Thresholds of hypoxia for marine biodiversity. *PNAS* 105(40), 15452–15457. <https://doi.org/10.1073/pnas.0803833105>
- Whitehead P.G., Wilby R.L., Battarbee R.W., Kernan M., and Wade A.J. (2009). A review of the potential impacts of climate change on surface water quality. *Hydrol. Sci. J. Sci. Hydrologiq*, 54(1),
- Whitford, L. A., and Schumacher, G. J. (1984). *A Manual of Freshwater Algae* (revised ed.). Sparks Press.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.
- Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *dplyr: A grammar of data manipulation*. R package version 1.1.4. <https://CRAN.R-project.org/package=dplyr>
- Work Economic Forum (WEF) (2023). *Sand mining is close to being an environmental crisis. Here's why – and what can be done about it*. Forum Institutional
- Wu, F., Huang, W., Liu, Q., Xu, X., Zeng, J., Cao, L., Hu, J., Xu, X., Gao, Y., and Jia, S. (2018). Responses of antioxidant defense and immune gene expression in early life stages of large yellow croaker (*Pseudosciaena crocea*) under methyl mercury exposure. *Front. Physiol.* 9, 1436. <https://doi.org/10.3389/fphys.2018.01436>
- Xaaceph Khan, M., and Butt, A. (2023). Evaluation of Water Quality Using Physicochemical Parameters and Aquatic Insects Diversity. In *River Basin Management - Under a Changing Climate*. IntechOpen. <https://doi.org/10.5772/intechopen.108423>
- Yaginuma, L.E, Gallucci, F., Vieira, D.C, Gheller, P.F., Brito de Jesus, S., Corbisier, T.N., and Fonseca, G. (2025) Hybrid machine learning algorithms accurately predict marine ecological communities. *Front. Mar. Sci.* 12, 1458014. <https://doi.org/10.3389/fmars.2025.1458014>
- Yang, Y., and Wang, H. (2025). Random Forest-Based Machine Failure Prediction: A Performance Comparison. *Applied Sciences*, 15(16), 8841. <https://doi.org/10.3390/app15168841>
- Yusuf, Z.H. (2019). Benthic macroinvertebrates diversity as bioindicators of water quality of Nasarawa Reservoir Katsina State Nigeria. *Bayero J. Pure Appl. Sci.*, 12, 449–456.

Journal currently Publish by the Publisher

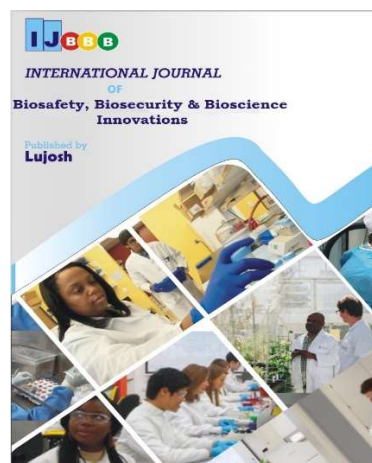
Kindly publish your research with us



Journal of Biotechnology
<https://www.jbt.com.ng>



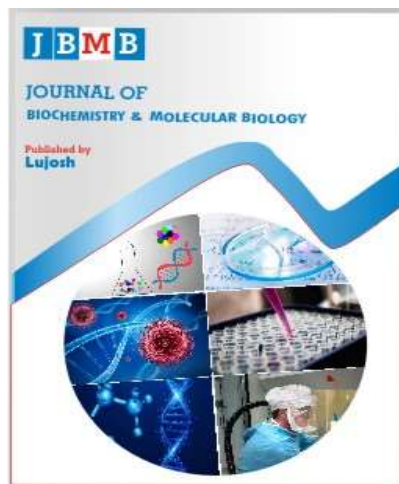
Environmental Studies Journal
<https://www.esj.com.ng>



**International Journal of Biosafety,
 Biosecurity and Bioscience**



**Journal of Agricultural
 Sciences and Technology**
<https://www.jast.com.ng>



**Journal of Biochemistry
 and Molecular Biology**
<https://www.jbmb.com.ng>



Plant Science Journal
<https://www.psjournal.com.ng>



**Journal of Parasites
Territory**
<https://www.jpt.com.ng>



Microbes World Journal
<https://www.mwj.com.ng>



Conference Proceedings
<https://www.conferenceproceedings.com.ng>