The Kumbotso Teacher Educator (KUTEJ):
A Journal of Multidisciplinary Studies
Institutional Based Research (IBR) SPECIAL EDITION
Vol 14.1 January 2025, p-ISSN: 1597-8796, e-ISSN: 2659-1995

Water Quality Prediction Model Based Support Vector Machine Model for Challawa Water Shed, Kano, Nigeria

Dr. Salisu Gwadabe, Muhammad Yunusa Adam and Hafsat Sani Ya'u
talktosalisugwadabe@gmail.com,

Department of Geography: Sa'adatu Rimi College of Education Kumbotso, Kano
Email:talkosalisugwadabe@gmail.com

Abstract

Water quality issues in Nigeria, exacerbated by industrial and urban expansion, pose significant health risks. A large portion of thepopulation relies on untreated surface water, leading to high incidences of waterborne diseases. In Kano, industrial effluents and agricultural runoff have further deteriorated water quality, as evidenced by pollution in the Challawa River. Despite some areas meeting standards for safe use, the need for continuous monitoring and predictive modelling is evident. This study aims to develop a Support Vector Machine (SVM) model for predicting water quality in the Challawa Watershed, addressing the limitations of current monitoring practices. The study involved a two-step sampling process based on land use, with 17 assessment locations identified during a reconnaissance survey and linked to their geographic coordinates using ArcGIS. Water samples were collected using clean plastic bottles, with field measurements of temperature, pH, and dissolved oxygen, while other parameters were analysed in the laboratory. The SVM model was trained on these data, with performance evaluated using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) to compare predicted values with actual measurements. The SVM model exhibited strong performance for predicting the pH and Potassium with low MAE and RMSE values. However, predictions for Chemical Oxygen Demand (COD) and Electrical Conductivity showed significant error, indicating reduced accuracy for these parameters. This variability underscores the model's strengths and limitations, highlighting areas for further improvement. While the SVM model is effective for certain water quality parameters, its performance is limited for complex indicators like COD and Electrical Conductivity. To enhance predictive accuracy, future research should refine the SVM model and incorporate additional data sources. Exploring alternative modelling techniques may also provide better results for challenging parameters and improve the overall water quality management.

Keywords: Challawa watershed; Industrial effluent; Pollutions; SVM, Water Quality,

Introduction

Water quality degradation in Nigeria poses a significant threat to environmental sustainability and public health, particularly in rapidly urbanizing and industrializing regions such as Kano State. The Challawa River, one of the major surface water sources in the region, has been the subject of growing concern due to increasing levels of pollution linked to industrial effluents and urban runoff. Numerous studies have highlighted the deteriorating water quality in this watershed, primarily due to uncontrolled discharges from the Challawa Industrial Estate (Faith et al., 2020; Abdulsalam, Mohammed & Kwajaffa, 2023).

A comprehensive physico-chemical analysis by Faith et al. (2020) identified high levels of heavy metals such as chromium (Cr^{2+}), lead (Pb^{2+}), and cadmium (Cd^{2+}), with chromium concentrations being the most dominant. The acidic nature of the water and varying dissolved oxygen levels further indicated anthropogenic stress on the ecosystem. Similarly, Abdulsalam et al. (2023) observed that although some water quality parameters such as Total Dissolved Solids (TDS) and Electrical Conductivity (EC) were within WHO permissible limits, effluent discharge still posed episodic contamination risks, particularly downstream of industrial outfalls. Mshelia et al. (2024) extended this concern to groundwater systems, revealing that boreholes in effluent-prone zones contained elevated concentrations of toxic metals like arsenic and mercury, surpassing World Health Organization (WHO) and Nigeria's NSDWA thresholds. These findings collectively highlight the pressing need for improved monitoring and predictive water quality management systems within the Challawa watershed.

Traditional monitoring approaches based on laboratory testing and periodic field sampling though reliable, often lack the responsiveness needed to detect and mitigate contamination in real-time. Consequently, the integration of **machine learning (ML)** techniques into water quality assessment has gained momentum as an effective alternative to manual methods (Zhu et al., 2022; Nasir et al., 2022). Unlike deterministic models, ML algorithms can detect complex nonlinear relationships in environmental datasets, offering robust tools for water quality classification, prediction, and trend analysis.

Recent advancements in ML applications to water resource management have demonstrated remarkable potential in predicting water quality indices (WQIs). Nasir et al. (2022) successfully utilized classifiers such as Support Vector Machine (SVM), CATBoost, and XGBoost to achieve over 94% accuracy in water quality prediction, while ensemble methods reached up to 100%. Similarly, Riaz et al. (2024) combined SVM with weight-of-evidence techniques, achieving over 99% accuracy in classifying water samples based on pH, TDS, and turbidity. In arid and industrialized regions, hybrid models such as Gaussian Process Regression (GPR) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been shown to be highly accurate in predicting Groundwater Quality Index (GWQI) and Water Pollution Index (Jibrin et al., 2024). These techniques not only improve prediction precision but also enable authorities to prioritize interventions by identifying key pollution drivers through feature sensitivity analyses (Mamat, Razali & Hamzah, 2023).

In the context of the Challawa River, the integration of ML models particularly SVM into water quality prediction presents a timely solution. Given the dynamic and multifactorial nature of pollution in this watershed, an SVM-based model can harness physico-chemical data to forecast pollution trends, identify high-risk zones, and provide early warnings. This study aims to develop a predictive water quality model for the Challawa River using SVM and relevant water quality parameters. By bridging field-based assessments with advanced machine learning techniques, the research will not only offer real-time decision support for environmental regulators but also contribute to broader efforts in sustainable water resource management across Nigeria.

Material and Method

Sampling

Sampling for surface water quality used a two-step process: Initially, dense sampling points were collected based on land use during a reconnaissance survey and linked to their geographic coordinates using ArcGIS. In the second step, after land use analysis, points were systematically subsampled to achieve 17 assessment locations, ensuring detailed coverage and summary conditions (Figure 1).

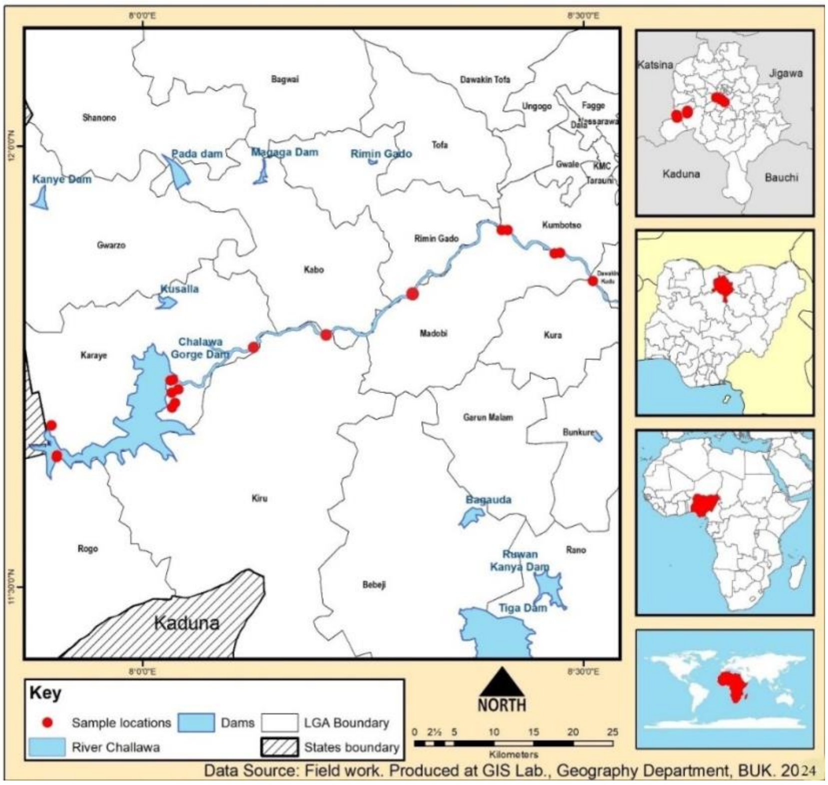

Figure 1: Distribution of Sample Point across the River
Source: GIS Lab, Geography Department, BUK, 2024

Sample Collection

Laboratory supplied plastic bottles were used for water sample collection. The equipment used for sample collection and field measurement were in good condition (they are cleaned periodically to prevent a build up of dirt). All field measurement instruments were fully calibrated before sampling collection could start (pre-field) and once again, all sampling have been completed (post-field). The results of the calibration were recorded in the field observation form. Surface water samples were directly collected along the river course supplying the watershed into clean plastic bottles in points shown in Figure 1. The collected surface water samples were kept in a cooler box containing iced-block to maintain a temperature of 1-4ºC which assisted in the preservation of the water sample. This procedure was repeated for all the water samples taken for four different periods.

Water Quality Analysis

Water quality parameters were assessed using a combination of field measurements and standardized laboratory techniques to ensure accuracy and reproducibility. Temperature, Dissolved Oxygen (DO), pH, and Electrical Conductivity (EC) were measured *in situ* using pre-calibrated handheld multi-parameter meters, which allow real-time analysis of these indicators directly at the sampling sites (APHA, 2017).

In the laboratory, Total Dissolved Solids (TDS) were determined using a conductivity probe. TDS values were calculated by applying a conversion factor to conductivity readings, accounting for the ionic composition of the water sample (EPA, 2012).

Chemical Oxygen Demand (COD) was analysed using the closed reflux colorimetric method, following the guidelines of the Standard Methods for the Examination of Water and Wastewater (APHA, 2017). This involved digestion of the water sample with potassium dichromate in a sulfuric acid medium, followed by spectrophotometric measurement.

Turbidity was measured using a nephelometric method with a calibrated turbidimeter, in accordance with EPA Method 180.1, which expresses results in Nephelometric Turbidity Units (NTU).

Nitrate (NO_3^-) and Sulphate (SO_4^{2-}) concentrations were determined using UV spectrophotometry and turbidimetric methods respectively, as described in APHA (2017). Chloride (Cl^-) levels were measured by argentometric titration using silver nitrate, with potassium chromate as an indicator.

Sodium (Na^+) and Potassium (K^+) were analysed using a flame photometer, a reliable method for detecting alkali metals in aqueous solutions (WHO, 2011). Biological Oxygen Demand (BOD_5) was determined using the 5-day incubation method at 20°C, measuring the oxygen consumed by microorganisms during organic matter decomposition (APHA, 2017).

Total Hardness and Alkalinity were quantified through titrimetric methods: hardness was measured using EDTA titration, while alkalinity was determined by titrating against a standard acid solution to the phenolphthalein and methyl orange endpoints (APHA, 2017).

All laboratory procedures were conducted under controlled conditions using quality-assured reagents and equipment. Calibration of instruments and quality control checks were performed regularly to maintain analytical precision and validity.

Data Analysis

The method of data analysis for the study on predicting water quality in the Challawa Watershed using a Support Vector Machine (SVM) involves several key steps to ensure a comprehensive evaluation of the model's performance. This approach includes descriptive statistics of water parameters, SVM-based predictions compared to actual data, and the evaluation of model accuracy through Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). Descriptive statistics provide an overview of the water quality parameters, while SVM plots of actual versus predicted values offer visual insights into the model's accuracy. MAE and RMSE serve as key metrics for evaluating the precision of the SVM model, ensuring that the predictions are both reliable and actionable. This comprehensive approach enables a thorough assessment of the model's effectiveness in predicting water quality in the Challawa Watershed.

Result

The analysis of water quality in the Challawa River revealed notable variations across different seasons. The dry season recorded the highest mean water temperature of 29.3°C, while the end of the rainy season followed at 28.6°C, indicating slight temperature fluctuations throughout the year. ANOVA tests confirmed significant differences in temperature across the watershed. The average DO levels were consistently below 7mg/l across all periods, with the highest variability observed during the dry (±10.1mg/l) and end of the rainy (±9.6mg/l) seasons. Despite this, statistical significance in the DO variation was not detected, possibly due to the minimal temperature differences. PH values ranged from 7.6 to 8, suggesting slightly alkaline conditions, which align with acceptable standards for both domestic and irrigation purposes (Wakawa *et al*., 2008; Boyi *et al*., 2017). TDS and COD were lowest during the dry season and highest at the end of the rainy season, with COD showing statistically significant variation. Turbidity peaked during the dry season, exceeding 100 NTU, though not significantly different across periods. Nitrate, Sulphate, Potassium, and Chloride levels were highest in the dry season, while Sodium and hardness were elevated during the rainy season. Biochemical Oxygen Demand (BOD) and alkalinity showed significant seasonal variations. Electrical conductivity remained consistent across seasons with no significant difference. Overall, turbidity was the only parameter exceeding WHO standards, reflecting similar findings from previous studies (Wakawa *et al*., 2008; Olanrewaju *et al*., 2017) and indicating increased pollution during the rainy season due to anthropogenic activities (Ioryue *et al*., 2018; Garizi *et al*., 2011).

Table 1 **Physicochemical Qualities of Water in Challawa River**

Parameters	**Dry Season**		**Rainy Season**		**Peak of Rainy Season**		**End of Rainy season**		**ANOVA**		**Standard Limit**
	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD	F	p-value	
Temp °c	29.3	1.9	27.2	0.8	27.8	0.9	28.6	1.6	6.99	**.000***	_
DO_2 mg/l	6.0	10.1	4.3	2.2	5.5	1.5	6.7	9.6	.372	.773	_
pH	7.8	0.8	7.6	0.6	8.0	0.8	7.6	0.6	1.187	.322	6.5-8.5
TDS mg/l	61.8	66.9	336.3	354.3	314.2	892.3	376.2	837.1	.603	.616	600-1000**
COD mg/l	199.3	262.4	568.1	1072.0	1896.6	2251.7	610.6	1397.7	3.889	**.013***	_
Turbidity	314.6	778.8	67.1	44.7	51.1	87.9	70.7	85.2	1.752	.166	10
Nitrate mg/l	16.5	14.9	4.3	3.3	1.1	2.5	57.0	123.6	2.684	.055	50
Sulphate mg/l	37.3	26.1	33.4	39.7	29.0	21.6	36.6	23.4	.269	.848	400
Chloride mg/l	9.1	7.5	1.9	1.9	5.4	5.3	3.9	3.8	5.657	**.002***	200
Sodium mg/l	327.9	300.4	457.1	503.8	154.2	166.6	42.9	22.5	6.00	**.001***	200-300
BOD mg/l	5.8	10.7	3.7	2.2	3.7	4.0	5.8	4.2	.731	.538	_
Hardness mg/l	1.3	0.8	4.7	2.7	4.0	1.8	0.5	0.8	21.278	**.000***	_
Alkalinity mg/l	0.7	0.5	1.0	0.7	101.5	78.9	42.0	78.0	10.876	**.000***	_
Potassium mg/l	2.5	1.9	2.4	0.7	2.1	1.5	1.3	1.5	2.335	.083	_
E/Cond	612.1	1701.3	476.9	1435.3	501.6	1621.1	669.4	1474.9	.057	.987	_

- **= No permissible standard limit**

Source: Analysis, 2024

The performance of SVM in predicting the parameters of water quality in Challawa is presented in figure 2 and 3. In Figure 2, actual data versus the predicted data were visualised while figure 3 shows the accuracy of the predicted values. The performance of the Support Vector Machine (SVM) model in predicting various water quality parameters demonstrates notable variability. For parameters such as pH and Potassium, the SVM model exhibits strong predictive capability with low Mean Absolute Error (MAE) values of 0.522 and 1.263, respectively, and Root Mean Square Error (RMSE) values of 0.628 and 1.600, respectively. This indicates that the SVM model can accurately estimate these parameters with relatively minimal error. In contrast, parameters like Turbidity and Nitrate show moderate performance, with MAE values of 34.948 and 9.033 and RMSE values of 41.320 and 12.347. While these results are higher compared to pH and Potassium, they still reflect a reasonable level of predictive accuracy. However, the model struggles significantly with predicting COD and Electrical Conductivity (E.Cond), evidenced by high MAE values of 821.248 and 389.431 and RMSE values of 1764.281 and 1133.458, respectively. This suggests that the model's accuracy is compromised for these parameters, indicating a need for further refinement or additional data.

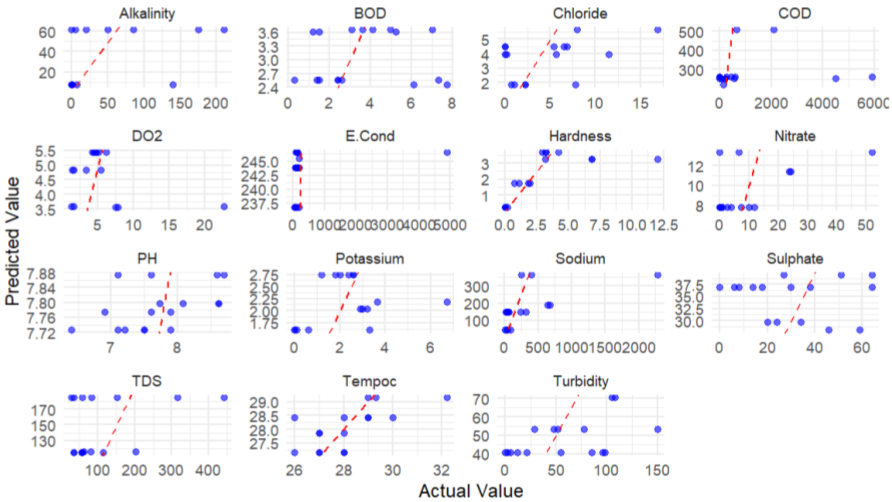

Figure 2: Scatter Plot of SVM Predicted Values Versus Actual Data

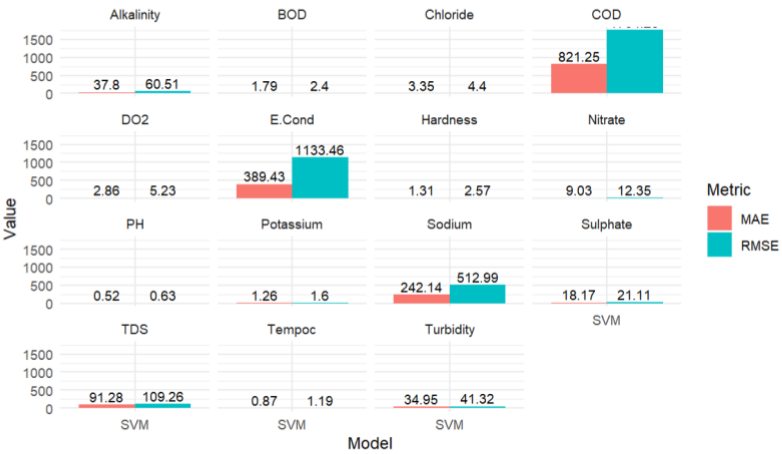

Figure 3: Performance of the Predictive Model

Discussion

The findings of this study align with and extend the insights from earlier research, particularly the work by Abobakr Yahya et al. (2019), which developed a SVM model to predict water quality parameters for ungauged river catchments. The study highlighted the effectiveness of SVM models in predicting various water quality indicators, including pH, Suspended Solids, Dissolved Oxygen, Ammonia Nitrogen, COD, and BOD. Yahya et al. reported high correlation coefficients (0.998 and 0.979) and low Mean Squared Error (MSE) values, demonstrating the model's accuracy in water quality prediction.

In contrast, this study observed notable variability in SVM performance across different parameters. For instance, the model achieved good predictive accuracy for pH and Potassium with low MAE and RMSE values, reflecting a robust performance similar to the results reported by Yahya et al. However, the study also found that the SVM model struggled with predicting COD and Electrical Conductivity, showing high MAE and RMSE values. This discrepancy suggests that while the SVM model is effective for some parameters, it faces challenges with others, indicating the need for further refinement or supplementary data to improve accuracy for complex parameters (Abobakr Yahya et al., 2019).

The implication of these findings for water quality management is significant. Effective prediction of key parameters like pH and Potassium can enhance water quality monitoring and management. However, the challenges in predicting COD and Electrical Conductivity highlight the need for advanced modelling techniques or additional data sources to address these parameters comprehensively. This study builds on the work of Yahya et al. by providing a nuanced view of SVM model performance, emphasizing the importance of parameter-specific adjustments and further research to enhance predictive accuracy across diverse water quality indicators.

Conclusion

This study successfully developed and evaluated a SVM model to predict water quality parameters within the Challawa Watershed in Kano, Nigeria. The research contributes significantly to the field of environmental monitoring by demonstrating the applicability of machine learning techniques specifically SVM in forecasting water quality in a region challenged by industrial pollution, agricultural runoff, and limited real-time monitoring infrastructure. The study's major contributions include providing a systematic assessment of water quality parameters across different seasons and validating the effectiveness of SVM in predicting parameters such as pH and Potassium with high accuracy and low prediction error.

Furthermore, the study advances current knowledge by identifying the spatial and temporal variability of pollution levels across the watershed, revealing that turbidity exceeded WHO standards during the dry season while parameters such as nitrate, sodium, and chloride exhibited critical seasonal fluctuations. By integrating GIS and machine learning, this research introduces a dynamic, data-driven framework that can support proactive water resource management, pollution mitigation, and early warning systems in Nigeria and other developing regions facing similar water quality challenges.

The findings highlight that while SVM is a valuable predictive tool for some indicators, its performance varies significantly across parameters, with limitations in predicting complex pollutants like COD and E.Cond. This underscores the importance of model-specific customization and the incorporation of additional environmental data for enhanced predictive performance.

Limitations and Recommendations for Future Study

One of the key limitations of this study is the SVM model's reduced predictive accuracy for certain complex parameters, particularly COD and Electrical Conductivity. These limitations may stem from the nonlinear nature of pollutant dynamics or insufficient representation of influential variables within the dataset. The study's reliance on available water quality measurements and the absence of real-time environmental variables (e.g., rainfall intensity, flow rate, land use changes) may have further constrained the model's accuracy.

To address these challenges, future studies should consider incorporating a broader range of input variables, including hydrological and meteorological data, and explore hybrid modelling approaches that combine SVM with other algorithms like Random Forests, Artificial Neural Networks, or Deep Learning methods. Additionally, increasing the frequency and spatial resolution of water quality sampling could help improve model generalization and parameter-specific calibration. Longitudinal studies that monitor changes over extended periods would also be beneficial for assessing the long-term applicability and robustness of machine learning-based prediction tools in water quality management.

By addressing these limitations and building upon the current framework, future research can enhance predictive capabilities, support evidence-based policymaking, and contribute to safeguarding public health and environmental sustainability in rapidly urbanizing and industrializing regions.

References

Abdulsalam, A., Mohammed, D., & Kwajaffa, H. M. (2023). The Influence of Industrial Effluent on Challawa River Water Quality. *FUDMA JOURNAL OF SCIENCES*, *7*(2), 294 - 299. https://doi.org/10.33003/fjs-2023-0702-2041

Abobakr Yahya, A.S.; Ahmed, A.N.; Binti Othman, F.; Ibrahim, R.K.; Afan, H.A.; El-Shafie, A.; Fai, C.M.; Hossain, M.S.; Ehteram, M.; Elshafie, (2019,) A. Water Quality Prediction Model Based Support Vector Machine Model for Ungauged River Catchment under Dual Scenarios. *11*, 1231. https://doi.org/10.3390/w11061231

APHA (2017). *Standard Methods for the Examination of Water and Wastewater*, 23rd Edition. American Public Health Association, American Water Works Association, Water Environment Federation.

Boyi, S., Yusuf, Y. O., Sawa, B. A., Adegbehin, A. B. (2017). An assessment of the Physicochemical Qualities of Water Sourcing in Kano Metropolis, Nigeria.

EPA (2012). *Method 180.1: Determination of Turbidity by Nephelometry*. United States Environmental Protection Agency.

Faith, O.C., Ali, A.F. and Odewade, L.O. (2020) "An Assessment of Water Quality Status of Challawa River in Kano State, Nigeria," *Dutse Journal of Pure and Applied Sciences (DUJOPAS)*, 6(1), pp. 68–76.

Garizi, Sheikh and Sadoddin (2011). Assessment of seasonal variations of chemical characteristics in surface water using multivariate statistical methods. *Int. J. Environ. Sci. Tech*., 8 (3),

Ioryue, I. S., Wuana, R. A., & Agustine, A. U. (2018). Seasonal Variation in Water Quality Parameters of River Mkomon Kwande Local Government Area, Nigeria. *International Journal of Recent Research in Physics and Chemical Sciences (IJRRPCS)* 5 (1). 42-62

Jibrin, A.M., Al-Suwaiyan, M., Aldrees, A. *et al. (2024)* Machine learning predictive insight of water pollution and groundwater quality in the Eastern Province of Saudi Arabia. *Sci Rep* **14**, 20031 (2024). https://doi.org/10.1038/s41598-024-70610-4

Mamat, N., Mohd Razali, S.F. and Hamzah, F.B. (2023) "Enhancement of water quality index prediction using support vector machine with sensitivity analysis," *Frontiers in environmental science*, 10. Available at: https://doi.org/10.3389/fenvs.2022.1061835.

Mshelia, S. Stephen , Oyetunji, B. Adewale , Ogar, P. Unimke and Reigns, A. N (2024). Appraisal of effects of industrial effluent on borehole water quality at Challawa industrial area in kano metropolis, Nigeria. *Water and Environmental Sustainability*, (),

Nasir, N. *et al.* (2022) "Water quality classification using machine learning algorithms," *Journal of water process engineering*, 48(102920), p. 102920. Available at: https://doi.org/10.1016/j.jwpe.2022.102920.

Olanrewaju A.N., Ajani E.K., Kareem O.K. & Orisasona O. (2017): Length-Weight Relationship and State of Well-being of Parachanna obscura (Gunther 1861) in Eleyele Reservoir, Southwestern Nigeria. F*isheries and Aquaculture Journal* 8: 3 DOI: 10.4172/2150-3508.1000221

Riaz, M.T., Riaz, M.T., Rehman, A. *et al. (2024)* An integrated approach of support vector machine (SVM) and weight of evidence (WOE) techniques to map groundwater potential and assess water quality. *Sci Rep* **14**, 26186 (2024). https://doi.org/10.1038/s41598-024-76607-3

Wakawa, R. J., Uzairu, A., Kagbu, J. A. and Balarabe, M. L. (2008). Impact assessment of effluent discharge on physico-chemical parameters and some heavy metal concentrations in surface water of River Challawa Kano, Nigeria. *African Journal of Pure and Applied Chemistry,* 2(9). 100-106. ISSN 1996 – 0840

WHO (2011). *Guidelines for Drinking-Water Quality*, 4th Edition. World Health Organization.

Zhu, M. *et al.* (2022) "A review of the application of machine learning in water quality evaluation," *Eco-Environment & Health*, 1(2), pp. 107–116. Available at: https://doi.org/10.1016/j.eehl.2022.06.001