ENGINEERING SCIENCE

The International Journal of Engineering Sciences
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

https://international-journal-of-engineering-sciences.jo

Machine Learning for Surface Water Quality Management in Malaysia: Integrating Physicochemical Parameters, Nutrient Loads, Land Use, Hydrological Conditions, and Anthropogenic Pressure

Iskandar M. Wahyu[1a,*] Gusti E. Rosyadi;[1b] May Abdul Hafed Abdul Kader;[2] Nuryani U. Humaira;[1c] Safaruddin P. Nasyita;[1d] Sudarijati;[1e] Ujang Asmil Zuwariah;[1f] Anisa Ricardi;[1g] Achuo Azmaine;[1h]

Extended author information available on the last page of the article**

* **Corresponding Author**: Iskandar M. Wahyu; iskandar_wahyu@usim.edu.my

DOI: https://doi.org/10.64440/IJES/EngineeringX0021

ARTICLE INFO

Article history
Received Apr 12, 2026
Revised Apr 26, 2026
Accepted Sep 10, 2026

Keywords
machine learning;
surface water quality;
Malaysia;
Water Quality Index;
Langat River Basin;
Random Forest;
Support Vector Machine;
LSTM;
land use;
nutrient pollution.

ABSTRACT

Surface water quality in Malaysia is increasingly influenced by rapid urbanization, industrial expansion, agricultural activities, wastewater discharge, climatic variability, and changes in hydrological regimes. These interacting pressures create complex and nonlinear relationships among physicochemical water-quality parameters, watershed characteristics, and pollutant dynamics, making conventional monitoring approaches insufficient for comprehensive and timely assessment. This study develops a data-driven machine learning framework for surface water quality assessment, prediction, pollution-event detection, and management in Malaysian river systems. The proposed framework integrates physicochemical, hydrological, spatial, land-use, and anthropogenic variables to characterize the temporal and spatial variability of water-quality conditions. The investigated parameters include dissolved oxygen, biochemical oxygen demand, chemical oxygen demand, pH, temperature, electrical conductivity, turbidity, total suspended solids, ammoniacal nitrogen, nitrate, phosphate, total nitrogen, and total phosphorus, together with rainfall, streamflow, water level, antecedent rainfall, urbanization, agricultural land use, industrial land use, population density, wastewater pressure, and road density. Seven machine learning approaches are incorporated, including Decision Trees (DT), Random Forest (RF), Support Vector Machines (SVM), Support Vector Regression (SVR), Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks, and hybrid ensemble models. The Malaysian National Water Quality Standards and Water Quality Index (WQI) are incorporated to establish the regulatory and quantitative basis for water-quality classification and prediction. The framework further integrates explainable artificial intelligence techniques, including SHAP and LIME, to quantify predictor contributions and identify the dominant environmental and anthropogenic factors governing model outputs. The proposed methodology enables simultaneous water-quality classification, continuous prediction, anomaly and pollution-event detection, source-pressure identification, and management-oriented decision support. Particular emphasis is placed on temporal variability, nonlinear interactions, hydrological influence, land-use effects, and the integration of multiple environmental data sources. The study provides a comprehensive Malaysia-oriented machine learning framework that can enhance conventional surface-water monitoring by combining predictive modeling, explainability, and integrated watershed information. The findings establish a methodological basis for developing real-time and transferable water-quality prediction systems capable of supporting pollution mitigation, watershed management, and sustainable water-resource planning across Malaysian river basins.

ENGINEERING SCIENCE
The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

1. Introduction

Surface water quality has been substantially affected during recent decades by the discharge of domestic wastewater [1], industrial effluents, agricultural pollutants, nutrients, suspended materials, organic contaminants, and other wastes [2]. Deteriorating water quality creates risks for human health, aquatic biodiversity, ecosystem functioning, water supply, irrigation, fisheries, and other beneficial uses [3,4,5]. The problem becomes more complex when climate variability modifies rainfall patterns, river discharge, water temperature, dilution capacity, and pollutant transport [6,7].

Sustainable surface-water management therefore requires systematic identification of the principal pollution sources and environmental pressures affecting rivers and other freshwater systems [8,9]. Traditional water-quality management relies heavily on periodic field sampling and laboratory analysis [10]. Although these procedures provide reliable measurements, their temporal and spatial coverage may be insufficient for rapidly changing pollution conditions [11]. Real-time sensors can improve temporal resolution, but their deployment can involve substantial acquisition, maintenance, calibration, and data-quality requirements [12].

Machine learning (ML) provides an alternative or complementary approach because it can identify nonlinear relationships among multiple environmental variables and transform large monitoring datasets into predictions, classifications, and early-warning information [13]. Previous research has demonstrated ML applications in water-quality classification, water-quality prediction, anomaly detection, pollution-source identification, and hydrological forecasting [14,15].

Malaysia provides a particularly relevant setting for this development. The Department of Environment (DOE) operates a national framework for river-water-quality standards and the River Water Quality Index [16]. Malaysian river waters are classified into Classes I, IIA, IIB, III, IV, and V according to designated uses and water-quality conditions [17]. Class I includes conservation and very limited-treatment uses, while Classes II–V correspond to progressively different treatment, fishery, recreational, livestock, irrigation, and other uses [18].

The Malaysian WQI framework also provides a direct basis for machine-learning classification. According to DOE documentation, WQI values of 81–100 are classified as clean, 60–80 as slightly polluted, and 0–59 as polluted [19]. The Malaysian classification framework incorporates parameters including ammoniacal nitrogen, biochemical oxygen demand, chemical oxygen demand, dissolved oxygen, pH, and total suspended solids [20].

Previous Malaysian research has already demonstrated the potential of ML for this purpose. Shamsuddin et al. developed ANN, Decision Tree, and Support Vector Machine models for multiclass water-quality classification in the Langat River Basin and reported that SVM provided strong classification performance [21]. Earlier research on the Langat River also combined chemometric techniques and ANN using extensive physicochemical measurements, identifying dissolved oxygen, BOD, pH, ammoniacal nitrogen, chlorine, E. coli, and coliform as important variables in explaining spatial water-quality variation [22].

However, a surface-water management system based only on conventional WQI parameters may not fully represent the processes controlling water quality. River condition is influenced not only by measured physicochemical properties but also by rainfall, discharge, catchment land use, agricultural activity, industrialization, urban expansion, wastewater inputs, and population pressure. Therefore, a more comprehensive ML framework should integrate these variables [23].

The present study develops such a framework for Malaysia. Rather than treating water quality as a function of water chemistry alone, the proposed framework conceptualizes surface-water quality as the product of interactions among five major dimensions: physicochemical conditions, nutrient loading, land use, hydrological conditions, and anthropogenic pressure.

The objectives are therefore to:

1. review the major ML algorithms applicable to surface-water-quality management;
2. establish a Malaysia-oriented set of physicochemical, nutrient, hydrological, land-use, and anthropogenic variables;
3. integrate Malaysian WQI and national water-quality classes into an ML decision framework;
4. evaluate the potential roles of DT, RF, SVM/SVR, ANN, LSTM, and ensemble models;
5. establish a framework for WQI classification, continuous water-quality prediction, and anomaly detection;
6. introduce explainable ML for identifying the environmental variables that most strongly influence model predictions; and
7. identify priorities for real-time, multi-source, and sustainable surface-water management in Malaysia.

2. Conceptual Framework for Surface Water Quality Management

2.1 Water Quality as a Multidimensional Environmental System

Water quality is a multidimensional concept involving physical, chemical, biological, and microbiological characteristics [24]. The conventional monitoring approach generally emphasizes direct measurements in water samples [25]. However, the observed value of a water-quality parameter can be influenced by hydrological conditions, upstream land use, pollution sources, meteorological conditions, and seasonal processes [26].

Consequently, the proposed ML framework represents surface-water quality through five interacting groups [27,28,29,30]:

Group 1: Physicochemical parameters

- pH
- dissolved oxygen (DO)
- water temperature (WT)
- electrical conductivity (EC)

ENGINEERING SCIENCE
The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

- turbidity
- total suspended solids (TSS)
- total dissolved solids (TDS)
- biochemical oxygen demand (BOD)
- chemical oxygen demand (COD)

Group 2: Nutrient parameters

- ammoniacal nitrogen (NH_3-N)
- nitrate (NO_3^-)
- total nitrogen (TN)
- phosphate
- total phosphorus (TP)

Group 3: Hydrological and meteorological parameters

- rainfall intensity
- cumulative rainfall
- antecedent rainfall
- streamflow/discharge
- water level
- seasonal conditions
- antecedent dry period

Group 4: Land-use parameters

- urban/built-up land
- agricultural land
- industrial land
- forest land
- impervious-surface proportion
- road density

Group 5: Anthropogenic pressure

- population density

- wastewater discharge
- industrial intensity
- agricultural intensity
- urban expansion
- proximity to pollution sources

This structure allows ML models to move beyond a purely chemical description of water quality and toward a catchment-scale representation of pollution processes.

2.2 Water Quality Index and Classification

The Water Quality Index (WQI) [31] provides a numerical representation of overall water condition by integrating multiple parameters into a single index.

A general weighted formulation can be expressed as:

$$WQI = \frac{\sum_{i=1}^{n} q_i w_i}{\sum_{i=1}^{n} w_i}$$

where:

- $q_i$is the sub-index value of parameter i;
- $w_i$is the corresponding parameter weight; and
- $n$is the number of parameters.

The conventional WQI approach is useful because it transforms multiple measurements into a common management indicator. However, the calculation requires measurements for multiple parameters and may become computationally demanding when monitoring frequency and spatial coverage increase.

ML can complement this process by learning the relationship between available input parameters and WQI or WQI class. The model can subsequently estimate water-quality condition from a reduced or expanded set of observations.

Table 1. Malaysian Water Quality Index Classification

WQI range	DOE classification	General interpretation
81–100	Clean	Relatively good water quality
60–80	Slightly polluted	Moderate deterioration
0–59	Polluted	Significant pollution

Source: Malaysian Department of Environment water-quality classification.

Malaysia additionally uses Classes I, IIA, IIB, III, IV, and V to link water-quality conditions with designated uses[32].

3. Machine Learning Algorithms for Surface Water Quality Management

ENGINEERING SCIENCE

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo

Vol. 2, No. 2, 2026, pp. 54-84

ISSN 3104-8897

3.1 Decision Tree

Decision Tree (DT) is a supervised-learning algorithm used for both classification and regression. It recursively divides observations according to selected predictor variables until terminal nodes representing predicted classes or numerical values are generated [33].

DT is particularly attractive for environmental applications because its decision structure can be interpreted comparatively easily. For example, a tree may identify combinations of low DO, high BOD, elevated ammoniacal nitrogen, and high TSS as conditions associated with degraded water-quality classes [34].

However, an individual tree can become excessively complex and overfit the training dataset. Pruning, cross-validation, and ensemble methods can therefore be used to improve generalization [35].

3.2 Random Forest

Random Forest (RF) combines multiple decision trees constructed using different bootstrap samples and subsets of predictor variables [36].

For water-quality management, RF has several advantages [37]:

1. ability to model nonlinear relationships;
2. capacity to accommodate many predictors;
3. relatively strong resistance to overfitting compared with individual trees;
4. ability to estimate variable importance;
5. suitability for classification and regression; and
6. applicability to mixed environmental datasets.

RF can therefore be used to predict WQI, classify water-quality classes, estimate individual pollutant concentrations, or identify the relative contribution of hydrological, land-use, and anthropogenic variables [38].

3.3 Support Vector Machine and Support Vector Regression

Support Vector Machine (SVM) is a supervised-learning method based on identifying an optimal decision boundary between classes. Kernel functions allow SVM to represent nonlinear relationships [39].

For continuous water-quality prediction, the corresponding Support Vector Regression (SVR) framework can be applied [40].

SVM/SVR is especially useful when [41]:

- datasets are relatively small;
- predictors are highly nonlinear;

- the number of variables is relatively high;
- generalization is important; and
- complex boundaries exist between water-quality classes.

A Malaysian study of the Langat River Basin demonstrated the applicability of SVM, ANN, and DT to multiclass WQI classification [42].

3.4 Artificial Neural Network

Artificial Neural Networks (ANNs) consist of interconnected computational nodes organized into input, hidden, and output layers [43].

For surface-water-quality management, ANN can learn nonlinear relationships between environmental variables and WQI or individual water-quality parameters [44].

For example:

$$WQI = f(D, O, B, O, D, C, O, D, p, HNH_3 - NTSSRainfallFlowLandUseAnthropogenicPressure)$$

The major advantage of ANN is its capacity to approximate highly nonlinear relationships [45]. However, ANN generally requires adequate training data, appropriate architecture selection, parameter tuning, and computational resources. Its internal decision process may also be difficult for environmental managers to interpret [46].

3.5 Long Short-Term Memory Networks

Water quality is inherently temporal. Measurements observed today may depend on conditions during previous hours, days, or weeks [47].

Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural network designed to learn temporal dependencies [48].

An LSTM-based framework can be represented conceptually as:

$$WQI_t = f(X_t, X_{t-1}, ., ..X_{t-k})$$

where $X_t$represents the environmental variables observed at time t, and $k$represents the historical time window.

Potential LSTM inputs include [49,50]:

- previous DO values;
- previous BOD/COD;
- rainfall;
- cumulative rainfall;
- streamflow;

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

- water level;
- nutrient concentrations;
- temperature; and
- previous WQI.

This makes LSTM particularly relevant for early-warning systems.

4. Proposed Malaysian Surface-Water ML Architecture

The proposed framework contains six sequential stages.

Stage 1: Multi-source data acquisition

Data should be collected from:

- DOE monitoring stations;
- automated water-quality sensors;
- meteorological stations;
- hydrological stations;
- GIS databases;
- satellite-derived land-use products;
- population databases;
- industrial databases; and
- wastewater-discharge records.

Stage 2: Data preprocessing

The dataset should undergo:

1. quality control;
2. missing-value treatment;
3. outlier identification;
4. temporal synchronization;
5. spatial matching;
6. normalization where required;
7. feature selection; and
8. class-balance assessment.

Stage 3: Feature integration

The five predictor groups should be merged into a unified feature matrix.

Stage 4: Model development

At least four model families should be compared:

- DT;
- RF;
- SVM/SVR;
- ANN/LSTM.

Ensemble models may subsequently combine the strongest models.

Stage 5: Model interpretation

SHAP and LIME should be used to identify the contribution of each predictor.

Stage 6: Management application

The final system should produce:

- WQI;
- water-quality class;
- pollutant predictions;
- anomaly alerts;
- pollution-pressure indicators; and
- management recommendations.

5. Proposed Variables and Their Environmental Roles

Table 2. Integrated predictor framework

Dimension	Variables	Expected management relevance
Physicochemical	pH, DO, WT, EC, turbidity, TSS, TDS	Direct representation of water condition
Organic pollution	BOD, COD	Organic and chemical pollution
Nutrients	NH_3-N, NO_3^-, TN, phosphate, TP	Nutrient enrichment and eutrophication
Hydrology	rainfall, discharge, water level	Pollutant dilution and transport
Climate	temperature, rainfall intensity, antecedent rainfall	Seasonal and event-driven changes
Land use	urban, agriculture, forest, industrial	Spatial pollution pressure

ENGINEERING SCIENCE

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo

Vol. 2, No. 2, 2026, pp. 54-84

ISSN 3104-8897

Urbanization	built-up area, impervious surface	Runoff and pollutant loading
Anthropogenic	population density, wastewater, industrial intensity	Human-generated pressure
Transport	road density, drainage connectivity	Pollutant mobilization
Management	WQI, WQC class	Final management indicators

The major methodological innovation of the proposed framework is therefore not simply replacing statistical models with ML. It is the integration of water chemistry with catchment-scale environmental drivers.

---

6. Machine Learning for Water-Quality Classification

Water-quality classification is one of the most established ML applications in surface-water management.

A classification model can estimate:

$$Y \in \{C, l, ean Slightly\ Polluted Polluted\}$$

or, at the Malaysian regulatory level:

$$Y \in \{I, I, I, A, I, IBIIIIVV\}$$

The input vector can contain both conventional water-quality parameters and contextual environmental variables.

Table 3. Candidate classification models

Model	Main application	Major advantage	Main limitation
DT	WQI classification	Easy interpretation	Overfitting
RF	Multiclass WQI	Robust nonlinear modeling	Lower transparency
SVM	WQI classification	Strong nonlinear classification	Parameter tuning
ANN	Complex classification	Nonlinear learning	Requires sufficient data
KNN	Similarity classification	Simple implementation	Sensitive to scale/outliers
XGBoost	Ensemble classification	High predictive capacity	Hyperparameter complexity
LSTM	Temporal classification	Learns temporal dependence	Data and computational demand

Existing Malaysian evidence supports the feasibility of this approach. Research in the Langat River Basin used ANN, DT, and SVM for multiclass water-quality classification, while other Malaysian work has investigated ANN-based WQI prediction.

---

7. Machine Learning for Continuous Water-Quality Prediction

Classification alone does not provide sufficient information for proactive management. Environmental agencies also require estimates of future pollutant concentrations and WQI.

A continuous prediction model can be formulated as:

$$\hat{Y}_{t+h} = f(X_t, X_{t-1}, \ldots X_{t-k})$$

where:

- $\hat{Y}_{t+h}$ is the predicted water-quality variable;
- h is the prediction horizon;
- X_t is the current environmental state; and
- k represents historical observations.

Potential prediction targets include:

- DO;
- BOD;
- COD;
- NH_3-N;
- NO_3^-;
- TP;
- TSS;
- turbidity;
- WQI; and
- water-quality class.

The inclusion of hydrological variables is particularly important because rainfall and discharge can alter pollutant concentration through dilution, mobilization, and transport.

8. Nutrient-Load Modeling

Nutrient pollution represents an important extension of conventional WQI-based modeling[51].

Nitrogen and phosphorus can originate from[52]:

- agricultural fertilizers;
- livestock operations;
- domestic wastewater;
- industrial wastewater;
- urban runoff;

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

- soil erosion; and
- atmospheric deposition.

The proposed ML framework therefore includes:

$$N_{load} = f(N، H_3، -، N، N، O_3^- TNRainfallFlowLandUse)$$

and

$$P_{load} = f(P، O_4^{3-}، T، P، RainfallFlowLandUse)$$

These relationships can be used to identify periods and locations where nutrient loading is likely to increase.

Importantly, concentration and load should not be treated as identical variables. Concentration describes pollutant abundance per unit water volume, whereas load incorporates flow and therefore provides information about the total pollutant mass transported through a river system[53].

9. Land Use and Spatial Water-Quality Modeling

Land use determines the spatial distribution of many pollution pressures[54].

Urban areas may contribute[55]:

- road runoff;
- suspended solids;
- metals;
- hydrocarbons;
- nutrients; and
- wastewater-related pressures.

Agricultural areas may contribute[56]:

- nitrogen;
- phosphorus;
- pesticides;
- sediments; and
- organic matter.

Industrial areas may contribute[57]:

- chemical pollutants;
- metals;

- organic contaminants; and
- elevated thermal or conductivity signals.

Forested areas generally represent a different hydrological and pollution-pressure regime[58].

Accordingly, GIS-derived land-use variables should be incorporated into ML models rather than treating every monitoring station as environmentally independent.

Potential spatial features include:

$$UrbanRatio = \frac{A_{urban}}{A_{catchment}}$$
$$AgriculturalRatio = \frac{A_{agriculture}}{A_{catchment}}$$
$$IndustrialRatio = \frac{A_{industrial}}{A_{catchment}}$$

and:

$$ImperviousRatio = \frac{A_{impervious}}{A_{catchment}}$$

These variables can be calculated for different buffer distances around monitoring stations or upstream subcatchments.

10. Hydrological Conditions and Water Quality

Hydrology provides an essential link between pollution sources and observed water quality[59].

Rainfall can increase surface runoff and transport pollutants from urban and agricultural surfaces into rivers. Conversely, prolonged low-flow conditions can reduce dilution capacity and increase the concentration of some pollutants[60].

The proposed framework therefore incorporates:

- current rainfall;
- cumulative rainfall;
- 24-h rainfall;
- 72-h rainfall;
- antecedent precipitation;
- discharge;
- water level;
- flow velocity where available; and
- dry-period duration.

ENGINEERING SCIENCE
The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

An antecedent rainfall index may be represented conceptually as:

$$API_t = P_t + kP_{t-1} + k^2P_{t-2} + \ldots + k^nP_{t-n}$$

where P_t is rainfall and k is a recession coefficient.

The inclusion of these variables allows the ML model to distinguish between pollution generated under normal flow and pollution mobilized during rainfall events.

11. Anthropogenic Pressure Index

A further extension of the framework is the construction of an Anthropogenic Pressure Index (API*) based on normalized human-pressure variables[61].

A general formulation can be expressed as:

$$API^* = \sum_{j=1}^{m} w_j Z_j$$

where Z_j represents a normalized anthropogenic-pressure variable and w_j represents its assigned or data-derived weight.

Potential components include:

- population density;
- industrial density;
- wastewater discharge;
- agricultural intensity;
- built-up proportion;
- road density; and
- proximity to wastewater outlets.

The purpose of this index is not to replace direct pollution measurements. Rather, it provides a contextual representation of human pressure that can improve spatial prediction and source interpretation.

12. Explainable Machine Learning

High predictive accuracy alone is insufficient for environmental decision-making. Managers need to know why a model predicts deterioration[62].

SHAP (SHapley Additive exPlanations) can quantify the contribution of individual variables to predictions[63].

For example, a SHAP analysis may reveal that[64,65]:

- rainfall has a strong positive contribution to TSS;

- high BOD contributes negatively to WQI;
- elevated DO contributes positively to WQI;
- urban land proportion increases pollution probability;
- ammoniacal nitrogen strongly influences WQI class.

LIME can provide local explanations for individual predictions.

The proposed framework therefore combines:

$$Prediction + Explanation + Environmental\ Interpretation$$

rather than prediction alone.

13. Water-Quality Anomaly Detection

Pollution incidents do not necessarily correspond to a gradual decline in WQI. Sudden changes can indicate[66,67]:

- accidental discharge;
- wastewater-system failure;
- intense storm runoff;
- illegal discharge;
- sensor malfunction; or
- unusual hydrological conditions.

Unsupervised ML can detect unusual observations without requiring a complete labeled database.

Potential methods include[68,69]:

- Isolation Forest;
- One-Class SVM;
- clustering;
- autoencoders;
- Bayesian approaches; and
- LSTM-based temporal anomaly detection.

An anomaly score can be used to generate early warnings:

$$A_t = f\left(X_t - \hat{X}_t\right)$$

where $X_t$represents observed conditions and $\hat{X}_t$represents expected conditions.

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

14. Performance Evaluation

The proposed models should be evaluated using metrics appropriate to the prediction task.

For regression:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - \hat{y}_i |$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

For classification:

- accuracy;
- precision;
- recall;
- F1-score;
- balanced accuracy;
- Cohen's kappa; and
- confusion matrix.

For imbalanced water-quality classes, accuracy alone should not be considered sufficient. Macro-F1 and class-specific recall should also be reported.

15. Validation Strategy

Random train-test splitting can produce overly optimistic results when environmental observations are temporally or spatially dependent.

Therefore, the Malaysian framework should use:

Temporal validation

Training on earlier observations and testing on later observations.

Spatial validation

Training on some monitoring stations and testing on geographically separate stations.

Catchment-level validation

Training on selected catchments and testing on an independent catchment.

Extreme-event validation

Testing specifically during high-rainfall and unusual pollution episodes.

These procedures provide stronger evidence of whether an ML model can generalize beyond the data used for model development.

16. Proposed Integrated Decision-Support Framework

Figure 1. Conceptual architecture of the proposed Malaysian ML system

Data sources

DOE monitoring + IoT sensors + rainfall + hydrology + GIS + land use + population + industrial/wastewater data

↓

Data preprocessing

Quality control → missing-value treatment → anomaly screening → normalization → temporal/spatial alignment

↓

Feature integration

Physicochemical + nutrient + hydrological + land-use + anthropogenic variables

↓

Machine-learning layer

DT | RF | SVM/SVR | ANN | LSTM | XGBoost | Hybrid models

↓

Outputs

WQI prediction + WQC classification + pollutant prediction + anomaly detection

↓

Explainability

SHAP + LIME + variable importance

↓

Management

Early warning + pollution-source prioritization + monitoring optimization + water-quality protection

17. Comparative Interpretation of Machine Learning Approaches

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

Table 4. Recommended role of ML models in the Malaysian framework

Model	Classification	Regression	Time series	Explainability	Recommended role
DT	High	High	Low	High	Transparent baseline
RF	High	High	Moderate	Moderate–High	Main ensemble model
SVM	High	High	Moderate	Moderate	Small/medium datasets
ANN	High	High	Moderate	Low	Nonlinear relationships
LSTM	Moderate	High	Very high	Low	Temporal forecasting
XGBoost	High	High	Moderate	Moderate–High	High-performance tabular data
KNN	Moderate	Moderate	Low	High	Benchmark/similarity
Hybrid models	High	High	Very high	Variable	Advanced prediction

18. Evidence from Malaysian Applications

The proposed framework is consistent with the direction of existing Malaysian research.

The Langat River Basin has been used extensively as a Malaysian case for data-driven water-quality analysis. Earlier research analyzed 23 physicochemical parameters from seven monitoring stations and used multivariate techniques and ANN to distinguish pollution regions[70,71,72,73,74,75,76,77,78,79].

More recent research developed ANN, DT, and SVM models for multiclass WQI classification in the Langat River Basin. The study showed that ML can reduce the computational complexity associated with conventional WQI classification and demonstrated the practical relevance of SVM-based classification.

Other Malaysian research has developed ANN models for WQI estimation in the Langat River, while ML studies have also investigated SVM-based water-quality prediction for catchments where direct observations may be limited.

These studies provide an empirical basis for applying ML to Malaysian river management. Nevertheless, the proposed framework extends the scope by integrating additional predictors representing land use, hydrology, nutrient loading, and anthropogenic pressure.

19. Advantages of the Proposed Framework

The integrated approach offers several potential advantages.

First, it can reduce dependence on a limited set of conventional water-quality variables by incorporating contextual environmental information.

Second, it can provide predictions between conventional sampling events.

Third, it can combine spatial and temporal information.

Fourth, it can identify variables associated with deterioration through feature-importance and explainability methods.

Fifth, it can support early-warning systems for sudden pollution events.

Sixth, it can improve monitoring efficiency by identifying locations and periods where additional sampling is most valuable.

Finally, it can provide a common analytical architecture that can potentially be transferred among Malaysian catchments after appropriate local calibration.

20. Limitations and Research Challenges

Despite its potential, several limitations must be addressed.

20.1 Data availability

ML performance depends heavily on the quantity, quality, temporal resolution, and representativeness of training data.

20.2 Missing observations

Environmental datasets frequently contain missing observations resulting from sensor failure, sampling interruptions, laboratory limitations, or changes in monitoring programs.

20.3 Spatial heterogeneity

Different Malaysian catchments have different hydrological, geological, climatic, land-use, and anthropogenic characteristics. A model trained in one basin should therefore not automatically be assumed to generalize to another.

20.4 Concept drift

Water-quality relationships can change because of urban expansion, climate variability, wastewater-treatment improvements, industrial changes, or land-use conversion.

20.5 Black-box models

Deep-learning models can provide strong predictive performance while remaining difficult to interpret.

20.6 Data leakage

Variables used to calculate WQI should be carefully distinguished from independent predictors when evaluating ML models intended to predict WQI. Otherwise, performance may be artificially inflated.

20.7 Regulatory compatibility

ML outputs should complement rather than replace official Malaysian regulatory procedures unless formally validated and adopted by the relevant authorities.

21. Future Research Directions

ENGINEERING SCIENCE
The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

Future Malaysian studies should focus on seven priorities.

21.1 Multi-source data fusion

The integration of DOE observations, IoT sensors, remote sensing, GIS, rainfall, hydrology, and socioeconomic data can produce a more comprehensive representation of catchment conditions.

21.2 Hybrid models

Hybrid architectures combining RF/XGBoost with LSTM or CNN-LSTM may be particularly useful for datasets containing both spatial and temporal patterns.

21.3 Explainable AI

SHAP, LIME, partial-dependence analysis, and feature-interaction analysis should become standard components of environmental ML studies.

21.4 Transfer learning

Models trained in data-rich Malaysian basins could potentially be adapted to data-limited basins through transfer-learning approaches, subject to rigorous validation.

21.5 Real-time monitoring

The integration of IoT sensors with ML could allow continuous water-quality assessment and automated warning generation.

21.6 Uncertainty quantification

Future models should provide prediction intervals rather than only point predictions.

21.7 Climate-change integration

Rainfall extremes, temperature changes, drought periods, and changing flow regimes should be incorporated into long-term water-quality prediction frameworks.

22. Management Implications for Malaysia

The proposed framework can support environmental management at several levels.

At the **monitoring level**, ML can identify which variables and stations require more frequent measurements.

At the **operational level**, real-time models can provide early warnings.

At the **catchment level**, land-use and anthropogenic variables can identify areas where pollution pressure is concentrated.

At the **regulatory level**, predicted WQI and water-quality classes can provide supplementary information for prioritizing investigation.

At the **strategic level**, long-term models can evaluate relationships between urbanization, climate variability, nutrient loading, and river-water quality.

Malaysia's existing DOE standards provide an appropriate regulatory reference because they link water-quality measurements to water-use classes and WQI categories.

23. Conclusions

Machine learning has substantial potential to strengthen surface-water-quality management in Malaysia by transforming conventional monitoring data into classification, prediction, anomaly-detection, and early-warning systems.

The principal contribution of the proposed framework is the integration of five dimensions of environmental information: **physicochemical parameters, nutrient loads, land use, hydrological conditions, and anthropogenic pressure**. This multidimensional approach recognizes that observed river-water quality is not determined exclusively by the chemical properties of a water sample but emerges from interactions among pollution sources, catchment characteristics, hydrological processes, climate conditions, and human activities.

DT, RF, SVM/SVR, ANN, LSTM, and ensemble algorithms each have different strengths. Tree-based methods provide strong performance on heterogeneous tabular datasets and can offer useful feature-importance information. SVM can be effective for nonlinear classification and relatively limited datasets. ANN can represent complex nonlinear relationships, while LSTM is particularly suitable for temporal forecasting. Hybrid models may provide additional advantages when spatial, temporal, and nonlinear processes interact.

The Malaysian regulatory context provides a strong foundation for ML development. The DOE framework uses WQI and water-use classes to characterize river-water quality, while existing Malaysian research has already demonstrated the feasibility of ANN, DT, SVM, and related approaches in the Langat River Basin.

The next generation of Malaysian water-quality ML systems should therefore move beyond models based solely on conventional WQI parameters. Integrating rainfall, streamflow, antecedent hydrological conditions, land-use composition, urbanization, agricultural activity, industrial pressure, population density, and wastewater information can provide a more complete representation of the mechanisms controlling surface-water quality.

At the same time, predictive accuracy should not be the only criterion for model selection. Explainability, uncertainty, spatial transferability, temporal robustness, data quality, regulatory compatibility, and operational feasibility are equally important. SHAP and LIME can help transform complex ML models into more interpretable decision-support tools by showing how individual environmental variables contribute to predictions.

A mature Malaysian surface-water ML system should ultimately connect **monitoring → prediction → explanation → early warning → management response**. Such an integrated architecture can complement conventional environmental monitoring, improve the efficiency of water-quality assessment, and provide a scientifically grounded foundation for proactive surface-water management.

The proposed framework is therefore suitable as a methodological basis for future empirical studies using Malaysian DOE monitoring records, IoT observations, hydrological datasets, remote-sensing products, and GIS-derived catchment characteristics. Its final empirical validation should be

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

performed using independently collected Malaysian observations and should include spatial, temporal, and event-based validation to establish the reliability and generalizability of the developed models.

Ethical Considerations

Not applicable. This study did not require ethical approval because it does not include human or animal subjects and does not involve any personal or sensitive data.

List of Abbrevations:

(WQI): The Water Quality Index ; (ML): Machine learning;(DOE): Department of Environment; (DO): dissolved oxygen; (WT): water temperature ; (EC): electrical conductivity; (TSS): total suspended solids ; (TDS): total dissolved solids; (BOD): biochemical oxygen demand; (COD): chemical oxygen demand; (NH_3-N): ammoniacal nitrogen; (NO_3^-): nitrate; (TN): total nitrogen; (TP): total phosphorus; (DT) : Decision Tree; (RF) : Random Forest;(SVM): Support Vector Machine ; (SVR) : Support Vector Regression; (ANNs): Artificial Neural Networks; (LSTM): Long Short-Term Memory.

Acknowledgment:

The authors would like to express their sincere gratitude to the **Faculty of Engineering and Technology, Multimedia University, Ayer Keroh 75450, Melaka, Malaysia**, for providing partial research funding under **Grant No. Malaysia.grant.2026.PTGF.26YH**, which supported the conduct of this study. The authors also gratefully acknowledge **The International Journal of Engineering Sciences – Noor Al-Ilm for Publishing and Distribution** for its generous support in providing a full waiver of the publication fees and facilitating the publication of this manuscript free of charge. The authors highly appreciate the institution's and journal's commitment to supporting scientific research and advancing opportunities for researchers.

Author Contribution:

All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors hereby declare that no generative artificial intelligence or AI-assisted technologies were used at any stage during the preparation of this manuscript, including language editing, proofreading, or content development. The authors take full responsibility for the originality and integrity of the work presented in this publication.

Funding:

This research received partial financial support from the **Faculty of Engineering and Technology, Multimedia University, Ayer Keroh 75450, Melaka, Malaysia**, under Grant No. **Malaysia.grant.2026.PTGF.26YH**, to support the conduct of the present study. The authors also acknowledge **The International Journal of Engineering Sciences, Noor Al-Ilm for Publishing and Distribution**, for providing a full waiver of the publication fees. The publication fee waiver was provided as editorial and publication support and did not involve any financial contribution to, or influence on, the research design, data collection, analysis, interpretation of results, or reporting of the research.

Conflicts of Interest:

"The authors declare no conflict of interest."

References

[1] Akhtar, N., Syakir Ishak, M. I., Bhawani, S. A., & Umar, K. (2021). Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review. Water, 13(19), 2660. https://doi.org/10.3390/w13192660

[2] Singh, P.K., Kumar, U., Kumar, I. et al. Critical review on toxic contaminants in surface water ecosystem: sources, monitoring, and its impact on human health. Environ Sci Pollut Res 31, 56428–56462 (2024). https://doi.org/10.1007/s11356-024-34932-0

[3] Ogidi, O.I., Akpan, U.M. (2022). Aquatic Biodiversity Loss: Impacts of Pollution and Anthropogenic Activities and Strategies for Conservation. In: Chibueze Izah, S. (eds) Biodiversity in Africa: Potentials, Threats and Conservation. Sustainable Development and Biodiversity, vol 29. Springer, Singapore. https://doi.org/10.1007/978-981-19-3326-4_16

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

[4] Reid, A.J., Carlson, A.K., Creed, I.F., Eliason, E.J., Gell, P.A., Johnson, P.T.J., Kidd, K.A., MacCormack, T.J., Olden, J.D., Ormerod, S.J., Smol, J.P., Taylor, W.W., Tockner, K., Vermaire, J.C., Dudgeon, D. and Cooke, S.J. (2019), Emerging threats and persistent conservation challenges for freshwater biodiversity. Biol Rev, 94: 849-873. https://doi.org/10.1111/brv.12480

[5] Zahoor, I., & Mushtaq, A. (2023). Water pollution from agricultural activities: A critical global review. Int. J. Chem. Biochem. Sci, 23(1), 164-176. https://iscientific.org/wp-content/uploads/2023/04/19-ijcbs-23-23-24.pdf

[6] WHITEHEAD, P. G., WILBY, R. L., BATTARBEE, R. W., KERNAN, M., & WADE, A. J. (2009). A review of the potential impacts of climate change on surface water quality. Hydrological Sciences Journal, 54(1), 101–123. https://doi.org/10.1623/hysj.54.1.101

[7] van Vliet, M.T.H., Thorslund, J., Strokal, M. et al. Global river water quality under climate change and hydroclimatic extremes. Nat Rev Earth Environ 4, 687–702 (2023). https://doi.org/10.1038/s43017-023-00472-3

[8] Akhtar, N., Syakir Ishak, M. I., Bhawani, S. A., & Umar, K. (2021). Various Natural and Anthropogenic Factors Responsible for Water Quality Degradation: A Review. Water, 13(19), 2660. https://doi.org/10.3390/w13192660

[9] Bănăduc, D., Simić, V., Cianfaglione, K., Barinova, S., Afanasyev, S., Öktener, A., McCall, G., Simić, S., & Curtean-Bănăduc, A. (2022). Freshwater as a Sustainable Resource and Generator of Secondary Resources in the 21st Century: Stressors, Threats, Risks, Management and Protection Strategies, and Conservation Approaches. International Journal of Environmental Research and Public Health, 19(24), 16570. https://doi.org/10.3390/ijerph192416570

[10] Arthur J. Horowitz; A Review of Selected Inorganic Surface Water Quality-Monitoring Practices: Are We Really Measuring What We Think, and If So, Are We Doing It Right?. Environ. Sci. Technol. 19 March 2013; 47 (6): 2471–2486. https://doi.org/10.1021/es304058q

[11] von Freyberg, J., Studer, B., and Kirchner, J. W.2017. A lab in the field: high-frequency analysis of water quality and stable isotopes in stream water and precipitation, Hydrol. Earth Syst. Sci., 21, 1721–1739, https://doi.org/10.5194/hess-21-1721-2017

[12] Skålvik AM, Saetre C, Frøysa K-E, Bjørk RN and Tengberg A (2023) Challenges, limitations, and measurement strategies to ensure data quality in deep-sea sensors. Front. Mar. Sci. 10:1152236. doi: https://doi.org/10.3389/fmars.2023.1152236

[13] Ferrario, D. M., Sanò, M., Maraschini, M., Critto, A., & Torresan, S. (2025). Harnessing machine learning methods for climate multi-hazard and multi-risk assessment. EGUsphere, 2025, 1-72. https://egusphere.copernicus.org/preprints/2025/egusphere-2025-670/

[14] Zou, S., Ju, H., & Zhang, J. (2025). Water Quality Management in the Age of AI: Applications, Challenges, and Prospects. Water, 17(11), 1641. https://doi.org/10.3390/w17111641

[15] Yu Li, Feng Han, Yi Zheng. Artificial Intelligence in Surface Water Quality Research and Management: Recent Progress and Future Directions. Ecosyst Health Sustain. 2026;12:0474.DOI: https://doi.org/10.34133/ehs.0474

[16] Al-Mamun, A., & Zainuddin, Z. (2013). Sustainable River Water Quality Management in Malaysia. IIUM Engineering Journal, 14(1). https://doi.org/10.31436/iiumej.v14i1.266

[17] Naubi, I., Zardari, N. H., Shirazi, S. M., Ibrahim, F. B. & Baloo, L. (2016). Effectiveness of Water Quality Index for Monitoring Malaysian River Water Quality. Polish Journal of Environmental Studies, 25(1), 231–239. https://doi.org/10.15244/pjoes/60109

[18] Zainudin, Z. (2010). Benchmarking river water quality in Malaysia. https://d1wqtxts1xzle7.cloudfront.net/71876031/Benchmarking_River_Water_Quality_in_Mala202110

08-25676-1xygf01.pdf?1738436245=&response-content-disposition=inline%3B+filename%3DBenchmarking_River_Water_Quality_in_Mala.pdf&Expires=1789894587&Signature=EnpO0He~rJli069zSRPYc4GN77cHgysb-66-hBrHXJPrhvZHqlPRVJP8NLVXt8eLIN1bF-n4168ktyLtO9Rhwbnil2uDN0Rpu10eFgGmgGRFLPkBclYMf7sacY2f8~3S9f3q0FGBf2XfZrgQSq7G45~EXidkdqxaIYpX1heiaj~PO2N7ZySJ0VVIkuSx5C56NnjS6yqxFsE16c-l2beSU5kXLzMR~124DilLHb~haE84tMkhNSUBjJV6cTrxpi-UdO0Kbxl5zC-cMeq5VeDWBNSO74PgtSLW6T3TccwWIQEME9M9cSy9FvfHvcx2gMPkp1H5Dj4YaKufPRsAyon8Tf__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[19] Malek, N. H. A., Wan Yaacob, W. F., Md Nasir, S. A., & Shaadan, N. (2022). Prediction of Water Quality Classification of the Kelantan River Basin, Malaysia, Using Machine Learning Techniques. Water, 14(7), 1067. https://doi.org/10.3390/w14071067

[20] Wen Yee Wong, Ayman Khallel Ibrahim Al-Ani, Khairunnisa Hasikin, Anis Salwa Mohd Khairuddin, Sarah Abdul Razak, Hanee Farzana Hizaddin, Mohd Istajib Mokhtar, Muhammad Mokhzaini Azizan,2022. Water Quality Index Using Modified Random Forest Technique: Assessing Novel Input Features, CMES - Computer Modeling in Engineering and Sciences, Volume 132, Issue 3, 2022, Pages 1011-1038, ISSN 1526-1492, https://doi.org/10.32604/cmes.2022.019244

[21] Shamsuddin, I. I. S., Othman, Z., & Sani, N. S. (2022). Water Quality Index Classification Based on Machine Learning: A Case from the Langat River Basin Model. Water, 14(19), 2939. https://doi.org/10.3390/w14192939

[22] Hafizan Juahir, Sharifuddin Md Zain, Ahmad Zaharin Aris, Mohd Kamil Yusoff, Mazlin Bin Mokhtar; Spatial assessment of Langat river water quality using chemometrics. J. Environ. Monit. 2010; 12 (1): 287–295. https://doi.org/10.1039/b907306j

[23] Banda, T. D., & Kumarasamy, M. (2024). Artificial Neural Network (ANN)-Based Water Quality Index (WQI) for Assessing Spatiotemporal Trends in Surface Water Quality—A Case Study of South African River Basins. Water, 16(11), 1485. https://doi.org/10.3390/w16111485

[24] Rocha, F.C., Andrade, E.M. & Lopes, F.B. Water quality index calculated from biological, physical and chemical attributes. Environ Monit Assess 187, 4163 (2015). https://doi.org/10.1007/s10661-014-4163-1

[25] Yolanda Madrid, Zoyne Pedrero Zayas, 2007. Water sampling: Traditional methods and new approaches in water sampling strategy, TrAC Trends in Analytical Chemistry, Volume 26, Issue 4, 2007, Pages 293-299, ISSN 0165-9936, https://doi.org/10.1016/j.trac.2007.01.002

[26] Tsegaye, T., Sheppard, D., Islam, K.R. et al. Development of Chemical Index as a Measure of In-Stream Water Quality in Response to Land-Use and Land Cover Changes. Water Air Soil Pollut 174, 161–179 (2006). https://doi.org/10.1007/s11270-006-9090-5

[27] Nayak, B., & Panda, P. K. (2026). A comprehensive review of water quality analysis. International Journal of Image and Graphics, 26(04), 2650033 https://www.worldscientific.com/doi/abs/10.1142/S0219467826500336

[28] Wang, Z. (2026). Data-Driven Approaches for Managing Water Quality in Urban Rivers (Doctoral dissertation, UNSW Sydney). https://unsworks.unsw.edu.au/entities/publication/f83b9fcb-0d57-47ea-bcb8-3f33a7c3a168

[29] Herano, D. Y., Woldesenbet, T. A., & Gebrekristos, S. T. (2025). Integrating principal component analysis and machine learning to assess riverine impacts on lake water quality: a case study of the Bilate River–Lake Abaya Watershed, Rift Valley, Ethiopia. Journal of Hydroinformatics, 27(12), 1838-1875. https://doi.org/10.2166/hydro.2025.085

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo

Vol. 2, No. 2, 2026, pp. 54-84

ISSN 3104-8897

[30] Wang, X. (2026). A Water Environment Monitoring System and Its Applications Based on Deep Learning. file:///C:/Users/DELL/Downloads/Xianhe%20Wang%20(1).pdf

[31] Banda, T. D., & Kumarasamy, M. V. (2020). Development of water quality indices (WQIs): A review. Polish Journal of Environmental Studies, 29(3), 2011-2021. https://www.pjoes.com/Development-of-Water-Quality-Indices-WQIs-nA-Review,110526,0,2.html

[32] Fazry, A. H. Z. (2022). Assessment of River Water Quality and Treats to Safe Water Supply in Malaysia (Master's thesis, University of Malaya (Malaysia)). https://www.proquest.com/openview/7a578ae9f460680f55e378c3f57a2614/1?pq-origsite=gscholar&cbl=2026366&diss=y

[33] Jadhav, P. V., & Patil, V. V. (2022). Application of Decision Tree for Developing Accurate Prediction Models. Ashok Yakkaldevi. https://books.google.jo/books?hl=en&lr=&id=b3N2EAAAQBAJ&oi=fnd&pg=PA1&dq=3.1+Decision+Tree+Decision+Tree+(DT)+is+a+supervised-learning+algorithm+used+for+both+classification+and+regression.+It+recursively+divides+observations+according+to+selected+predictor+variables+until+terminal+nodes+representing+predicted+classes+or+numerical+values+are+generated.&ots=-Picpk4p_n&sig=qNQNu6jt2feL42A7tvoHlswuFz0&redir_esc=y#v=onepage&q&f=false

[34] Anmala, J., & Turuganti, V. (2021). Comparison of the performance of decision tree (DT) algorithms and extreme learning machine (ELM) model in the prediction of water quality of the Upper Green River watershed. Water Environment Research, 93(11), 2360–2373. https://doi.org/10.1002/wer.1642

[35] Das, A. An optimization based framework for water quality assessment and pollution source apportionment employing GIS and machine learning techniques for smart surface water governance. Discov Environ 3, 117 (2025). https://doi.org/10.1007/s44274-025-00327-2

[36] Random Forest Algorithm Overview (H. A. Salman, A. Kalakech, & A. Steiti, Trans.). (2024). Babylonian Journal of Machine Learning, 2024, 69-79. https://doi.org/10.58496/BJML/2024/007

[37] Alnahit, A.O., Mishra, A.K. & Khan, A.A. Stream water quality prediction using boosted regression tree and random forest models. Stoch Environ Res Risk Assess 36, 2661–2680 (2022). https://doi.org/10.1007/s00477-021-02152-4

[38] Zhang, H., Ren, X., Chen, S., Xie, G., Hu, Y., Gao, D., ... & Wang, H. (2024). Deep optimization of water quality index and positive matrix factorization models for water quality evaluation and pollution source apportionment using a random forest model. Environmental Pollution, 347, 123771. https://www.sciencedirect.com/science/article/abs/pii/S0269749124004858

[39] Ukil, A. (2007). Support Vector Machine. In: Intelligent Systems and Signal Processing in Power Engineering. Power Systems. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-73170-2_4

[40] Yang, H., Zou, T., Huang, Y. et al. Spatiotemporal evolution of water quality in long-distance water supply projects: an improved PSO-SVR model. Environ Monit Assess 197, 1351 (2025). https://doi.org/10.1007/s10661-025-14805-w

[41] Xie, R., Zhou, Z., Kong, J., Wang, C., Wang, Y., Li, L., Ding, C., Li, R., & Zhang, X. (2025). Multi-Algorithm Comparison for Water Quality Retrieval: Integrating Landsat-8 OLI and Machine Learning in Karst Plateau Reservoirs. Water, 17(12), 1781. https://doi.org/10.3390/w17121781

[42] Shamsuddin, I. I. S., Othman, Z., & Sani, N. S. (2022). Water Quality Index Classification Based on Machine Learning: A Case from the Langat River Basin Model. Water, 14(19), 2939. https://doi.org/10.3390/w14192939

[43] Roza Dastres, Mohsen Soori. Artificial Neural Network Systems. International Journal of Imaging and Robotics (IJIR), 2021, 21 (2), pp.13-25. https://hal.science/hal-03349542/

[44] Zhang, Q., You, Xy. Recent Advances in Surface Water Quality Prediction Using Artificial Intelligence Models. Water Resour Manage 38, 235–250 (2024). https://doi.org/10.1007/s11269-023-03666-y

[45] Tu, J. V. (1996). Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. Journal of clinical epidemiology, 49(11), 1225-1231. https://doi.org/10.1016/S0895-4356(96)00002-9

[46] Oyebode O, Stretch D. Neural network modeling of hydrological systems: A review of implementation techniques. Natural Resource Modeling. 2019; 32:e12189. https://doi.org/10.1111/nrm.12189

[47] Cassidy, R., & Jordan, P. (2011). Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology, 405(1-2), 182-193. https://www.sciencedirect.com/science/article/abs/pii/S0022169411003301

[48] Yong Yu, Xiaosheng Si, Changhua Hu, Jianxun Zhang; A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Comput 2019; 31 (7): 1235–1270. doi: https://doi.org/10.1162/neco_a_01199

[49] Pan, D., Deng, Y., Yang, S. X., & Gharabaghi, B. (2025). Recent Advances in Remote Sensing and Artificial Intelligence for River Water Quality Forecasting: A Review. Environments, 12(5), 158. https://doi.org/10.3390/environments12050158

[50] Alaka, S. (2025). Integrating machine learning and IoT technologies for smart water quality monitoring: Methods, challenges, and future directions. Int J Data Sci Big Data Anal, 5(2), 46-68. https://www.preprints.org/frontend/manuscript/0bc170810a73179089f95c74bcbd254e/download_pub

[51] Mo, Y., Xu, J., Liu, C. et al. Assessment and prediction of Water Quality Index (WQI) by seasonal key water parameters in a coastal city: application of machine learning models. Environ Monit Assess 196, 1008 (2024). https://doi.org/10.1007/s10661-024-13209-6

[52] Xue, J., Yuan, C., Ji, X., & Zhang, M. (2024). Predictive modeling of nitrogen and phosphorus concentrations in rivers using a machine learning framework: A case study in an urban-rural transitional area in Wenzhou China. Science of the Total Environment, 910, 168521. https://www.sciencedirect.com/science/article/pii/S0048969723071498

[53] Bertrand-Krajewski, J. L., Chebbo, G., & Saget, A. (1998). Distribution of pollutant mass vs volume in stormwater discharges and the first flush phenomenon. Water research, 32(8), 2341-2356. https://doi.org/10.1016/S0043-1354(97)00420-X

[54] Heathwaite, A. L., Quinn, P. F., & Hewett, C. J. M. (2005). Modelling and managing critical source areas of diffuse pollution from agricultural land using flow connectivity simulation. Journal of Hydrology, 304(1-4), 446-461. https://doi.org/10.1016/j.jhydrol.2004.07.043

[55] Tariq, A., & Mushtaq, A. (2023). Untreated wastewater reasons and causes: A review of most affected areas and cities. Int. J. Chem. Biochem. Sci, 23(1), 121-143. https://iscientific.org/wp-content/uploads/2023/04/15-ijcbs-23-23-22.pdf

[56] Rashmi, I. et al. (2022). Soil Erosion and Sediments: A Source of Contamination and Impact on Agriculture Productivity. In: Naeem, M., Bremont, J.F.J., Ansari, A.A., Gill, S.S. (eds) Agrochemicals in Soil and Environment. Springer, Singapore. https://doi.org/10.1007/978-981-16-9310-6_14

[57] Saxena, V. Water Quality, Air Pollution, and Climate Change: Investigating the Environmental Impacts of Industrialization and Urbanization. Water Air Soil Pollut 236, 73 (2025). https://doi.org/10.1007/s11270-024-07702-4

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

[58] Ding, Q., Xi, H., Li, P., Fang, H., Yuan, Y., & Li, T. (2026). Linking water quality assessment to source apportionment with machine learning-assisted WQI, PMF, and SOM: A case study of the Jinma River basin. Ecological Informatics, 103625. https://doi.org/10.1016/j.ecoinf.2026.103625

[59] Brooks, B.W., Riley, T.M. & Taylor, R.D. Water Quality of Effluent-dominated Ecosystems: Ecotoxicological, Hydrological, and Management Considerations. Hydrobiologia 556, 365–379 (2006). https://doi.org/10.1007/s10750-004-0189-7

[60] Ferreira, C. S., Kalantari, Z., Seifollahi-Aghmiuni, S., Ghajarnia, N., Rahmati, O., & Solomun, M. K. (2021). Rainfall-runoff-erosion processes in urban areas. In Precipitation (pp. 481-498). Elsevier. https://doi.org/10.1016/B978-0-12-822699-5.00018-5

[61] Cegielska, K., & Salata, T. (2023). A multimetric approach to anthropogenic pressure analyses: A tool for measuring and mapping the human impact on the landscape. Land Degradation & Development, 34(17), 5293–5311. https://doi.org/10.1002/ldr.4845

[62] Infant, S. S., Vickram, S., Saravanan, A., Muthu, C. M., & Yuarajan, D. (2025). Explainable artificial intelligence for sustainable urban water systems engineering. Results in Engineering, 25, 104349. https://doi.org/10.1016/j.rineng.2025.104349

[63] García, M. V., & Aznarte, J. L. (2020). Shapley additive explanations for NO2 forecasting. Ecological Informatics, 56, 101039. https://doi.org/10.1016/j.ecoinf.2019.101039

[64] Towfiqul Islam, A.R.M., Mamun, M.AA., Uddin, M.N. et al. Explainable and physics-informed machine learning for seasonal water quality prediction in the monsoon-driven Padma River Basin, Bangladesh. Environ Geochem Health 48, 140 (2026). https://doi.org/10.1007/s10653-026-03031-z

[65] Fangnan Xiao, Runqiao Zhang, Zhanqiang Jian, Wei Liu, Taotao Sun, Weicheng Pang, Long Han, Huapeng Qin,2025. Using ensemble machine learning to predict and understand spatiotemporal water quality variations across diverse watersheds in coastal urbanized areas, Ecological Indicators, Volume 178, 2025, 113976, ISSN 1470-160X, https://doi.org/10.1016/j.ecolind.2025.113976

[66] Musaazi, I. G. (2025). Evaluating the Impact of Extreme Wet Weather Events on Biological Wastewater Processes Using Modeling (Doctoral dissertation, Duke University). https://www.proquest.com/openview/694c4b2da2984c9cf5e78da411990ded/1?pq-origsite=gscholar&cbl=18750&diss=y

[67] Payum, T. (2020). Advances in Environmental Pollution Management: Wastewater Impacts and Treatment Technologies. Advances in Environmental Pollution Management: Wastewater Impacts and Treatment Technologies. https://d1wqtxts1xzle7.cloudfront.net/109737433/aepm-2020-libre.pdf?1703845903=&response-content-disposition=inline%3B+filename%3DAdvances_in_Environmental_Pollution_Mana.pdf&Expires=1789900596&Signature=bfiWzNfSAQJhUfMIv8Qt5BHulGqKft9BWpD-ycFafczdQXCMv~iy1gBRK-2QF4Ma8KwPaKzVyb1POxNpcVAcpNX1Gbz3ZgmuSDH-zqVIDbQOr3OXD3TdmaZKHmlglCAxhsmHg~KKs01GaloSz04a9e4c57HkbY05f1sFU-WLM2qgJslN34Q87Bmkeg3x3hOe-EME9M9cSy9FvfHvcx2gMPkp1H5Dj4YaKufPRsAyon8Tf-gLC9X7q0yH8QNPjGfGh5Bki9XzBcHzdOEAEpnI~iqoMKZkdfOMWEYS6W0N6q0uqbWOLNvao1uel2ZOVAo-RmuO9UtPe0nlbF-6jc6B6bg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[68] K. Choi, J. Yi, C. Park and S. Yoon, "Deep Learning for Anomaly Detection in Time-Series Data: Review, Analysis, and Guidelines," in IEEE Access, vol. 9, pp. 120043-120065, 2021, doi: https://doi.org/10.1109/ACCESS.2021.3107975

[69] Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. ACM Computing Surveys, 57(1), 1-42. https://doi.org/10.1145/3691338

[70] Mamat N, Mohd Razali SF and Hamzah FB (2023) Enhancement of water quality index prediction using support vector machine with sensitivity analysis. Front. Environ. Sci. 10:1061835. doi: https://doi.org/10.3389/fenvs.2022.1061835

[71] Mamat, N., & Razali, S. M. (2023). Comparisons of various imputation methods for incomplete water quality data: a case study of the langat river, Malaysia. Jurnal Kejuruteraan, 35(1), 191-201. https://d1wqtxts1xzle7.cloudfront.net/120698992/jkukm-2023-351-18-libre.pdf?1736319841=&response-content-disposition=inline%3B+filename%3DComparisons_of_Various_Imputation_Method.pdf&Expires=1789901226&Signature=YH723oZN1sv0EpGIcCzMKnsku6f9RNlBnZd-4txFveLOEtYcZGOD2Sgv-6fgtzAUYnPZyohn4C0gl9Jen7y2uwTXzww4J5rAz1RadSjjJxSKVRtO22o6i-3qaef5auSfdHZ4iZnuCmk~SpoaHm4505jETJVUb38gTWgm2-4uUGAk0nHlzhtzJAO885Pp~hzsrRYfNnGwjJhunALLJ40eimr1B8OA-h-4L~IsvFlDs3YLfKA20JDEhVh2dj06SRVoE25bE~gr-uN3ip8O4cA0fJEJgQCZ1JcsAEBKiaSZwXEME9M9cSy9FvfHvcx2gMPkp1H5Dj4YaKufPRsAyon8Tf__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[72] Hameed, M. M., Kareem, A. S., Aljumaily, M. M., Masood, A., Afan, H. A., Al Omar, M. K., & Razali, S. F. M. (2023, December). Data-driven models for predicting water quality index: A case study of Klang and Langat rivers, Malaysia. In AIP Conference Proceedings (Vol. 3015, No. 1, p. 020017). AIP Publishing LLC. https://doi.org/10.1063/5.0188646

[73] Zainol, N. F. M., Zainuddin, A. H., Looi, L. J., Aris, A. Z., Isa, N. M., Sefie, A., & Ku Yusof, K. M. K. (2021). Spatial Analysis of Groundwater Hydrochemistry through Integrated Multivariate Analysis: A Case Study in the Urbanized Langat Basin, Malaysia. International Journal of Environmental Research and Public Health, 18(11), 5733. https://doi.org/10.3390/ijerph18115733

[74] Lim, F. H., Lee, W. K., Khor, W. S., Chan, L., Sulaiman, M. S., Shaari, H., & Jalil, M. K. (2023). River basin change index for integrated river basin management of Langat River, Malaysia. Water Policy, 25(11), 1035-1061. https://doi.org/10.2166/wp.2023.035

[75] Rizal, N. N. M., Hayder, G., & Yusof, K. A. (2022). Water Quality Predictive Analytics Using an Artificial Neural Network with a Graphical User Interface. Water 2022, 14, 1221. https://doi.org/10.3390/w14081221

[76] Shaheed, H., Zawawi, M. H., Hayder, G., Mostafa, K. S., Sidek, N., & Hafez, M. A. (2025). Comparative Analysis of River Water Quality in Asia and the Arab Region Using Machine Learning-Based Classification and Prediction Models. Applied Mathematics and Information Sciences, 19(6), 1501-1524. http://dx.doi.org/10.18576/amis/190620

[77] Abdullah, N.A.A., Juahir, H. Enhancing river water quality modelling using a PCA-based multivariate Water Quality Index with machine learning prediction across 14 Malaysian river basins. Model. Earth Syst. Environ. 12, 234 (2026). https://doi.org/10.1007/s40808-026-02871-z

[78] Azni, Q.I.M., Shazili, F.S., Serri, N.A. (2026). Fertilizer Runoff Management—Strategies and Practices to Minimise Excess Nutrients in Water Reservoirs. In: Yaser, A.Z., Rajin, M., Saalah, S. (eds) Compost and Liquid Organic Fertilizers. Springer, Singapore. https://doi.org/10.1007/978-981-95-8561-8_22

[79] Naufal Raja Azman, R. M., Mohd Noor, N. A., Ideris, M. M., & Abdullah, S. (2026). Assessing the Drought Index and Water Scarcity Nexus: Case Study for Langat River Basin. Pertanika Journal of Science & Technology, 34(4), 2213. https://openurl.ebsco.com/EPDB%3Agcd%3A12%3A40655372/detailv2?sid=ebsco%3Aplink%3Acrawler-gcd&id=ebsco%3Agcd%3A196671055&crl=c&jrnl=01287680&link_origin=scholar.google.com

The International Journal of Engineering Sciences
https://international-journal-of-engineering-sciences.jo
Vol. 2, No. 2, 2026, pp. 54-84
ISSN 3104-8897

Authors and Affiliations

Iskandar M. Wahyu[1a,*] Gusti E. Rosyadi;[1b] May Abdul Hafed Abdul Kader;[2] Nuryani U. Humaira;[1c] Safaruddin P. Nasyita;[1d] Sudarijati;[1e] Ujang Asmil Zuwariah;[1f] Anisa Ricardi;[1g] Achuo Azmaine;[1h]

[1] Faculty of Engineering and Technology, Multimedia University, Ayer Keroh 75450, Melaka, Malaysia

iskandar_wahyu@usim.edu.my	a	gusti.e.ros8@usim.edu.my	,b
u.humaira43@usim.edu.my	c	safar.p.nas@usim.edu.my	d
sudarijati1986@usim.edu.my	,e	ujanh_asm5@usim.edu.my	,F
anisa.ricardi9@usim.edu.my	g	achuo_azmaine@usim.edu.my	h

[2] Assistant Professor; Environmental Research Center; University of Mosul; Mosul-Iraq; Email: may_allaf@uomosul.edu.iq ; Orcid: https://orcid.org/0009-0005-1624-8128

*** Corresponding Author**: Iskandar M. Wahyu; iskandar_wahyu@usim.edu.my