

RESEARCH ARTICLE

Digital and Data Mining Approaches for Evaluating Water Pollution

ASIFAT Janet Temitope, OMIGBODUN Olanrewaju Tolu

University of Ilesa, Ilesa, Nigeria.

Received: 23 February 2026 Accepted: 09 March 2026 Published: 10 March 2026

Corresponding Author: ASIFAT Janet Temitope, University of Ilesa, Ilesa, Nigeria.

Abstract

The importance of digital and social media data for evaluating water pollution cannot be overemphasised. It enables the water pollution forecasts, hence the quality of data of information gathered from social media platforms must be maintained, especially for desert regions with intermittent but heavy rainfall and a significant risk of polluted areas. This study explored the digital and data mining, and social media data mining to evaluate water pollution. It was revealed that earlier studies concentrated more on creating a framework for social media activity in water pollution, especially in places that flood. This study examines reports and records of water pollution from a variety of social media sites, including YouTube, Flickr, Facebook, Instagram, and Twitter, for the first time. It verified data on water pollution from social media. Based on the results of many machine learning classifiers, social media data was validated to predict and evaluate water pollution.

Keywords: Data Mining, Digital Data Mining, Social Media Data Mining, Water Pollution.

1. Introduction

Hydrological research offers a lot of promise to improve water pollution control by using data on water pollution collected from social media (Adeniran, Sidiq *et al.*, 2024; Adeniran, Oyeniran *et al.*, 2024). When combined with other technical data from the watershed, photos of several water samples collected at different periods and with time stamps and geographic references can be used to forecast how the water pollution hydrograph would grow. Water pollution has negative socio-cultural and environmental effects on the people (Akinloye, 2024a; Akinloye, 2024b; Akinloye, 2025). Videos can depict the start of a flash flood, polluted water, and the passage of the water wave through the catchment in addition to the movement of various waste particles that define the pollution level. Models of flow can be calibrated with the use of such data.

Recently, several digital approaches have been used to evaluate water pollution; these techniques are useful for figuring out how pollutants have changed over time in terms of both space and time (Familusi, Omoyeni *et al.*, 2024; Yadav *et al.*, 2021). It is possible to mimic the spatiotemporal development of pollutants in “river” aquatic ecosystems using a variety of dynamic simulation software tools (Adeniran, Adeniran & Udorah, 2025). Numerous US companies have already produced this kind of software, including the Institute of Fresh water Ecology and Inland Fisheries in Berlin, Germany (MONERIS programme), the US Army Hydrologic Engineering Centre (WQRRS programme), the US Environmental Protection Agency (WASP, QUAL2E programmes), the US Ecology Division, Office of Research and Development (AQUATOX programme), ANSYS company from Canonsburg, Pennsylvania, USA

Citation: ASIFAT Janet Temitope, OMIGBODUN Olanrewaju Tolu. Digital and Data Mining Approaches for Evaluating Water Pollution. Annals of Ecology and Environmental Science. 2026; 8(1): 18-30.

©The Author(s) 2026. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

(ANSYS CFX programme), MapTech Company in Blacksburg, Virginia, USA (GWLF programme), the Institute of Freshwater Ecology and Inland Fisheries, Berlin, Germany (MONERIS programme), the USA company Aquaveo (WMS, SMS programmes), the US Army Hydrologic Engineering Centre (WQRRS programme), etc. (Nina & Galina, 2021; Tiyasha *et al.*, 2021).

2. Literature Review

The use of models or software packages has highlighted many previously unresolved scientific issues. This is a particularly flexible tool to adapt to other conditions through modification and improvement (Bui *et al.*, 2020). With their help, it is possible to reconstruct and perform simulations of the processes to which chemical compounds have been subjected in the past, and also to predict the future cycle of chemicals through different ecosystems both globally and regionally and for different time intervals (Ardabili *et al.*, 2019).

2.1 WASP

This programme, which is an enhancement of the original WASP, simulates water quality analysis (Eddine *et al.*, 2023). For a variety of pollution management choices, this model assists users in interpreting and forecasting reactions to water quality, natural events, and man-made pollution (Babbar & Babbar, 2017). For aquatic systems, including the water column and the underlying benthos, it is a dynamic compartment modelling programme (Tuan, 2020). It lets the user look at a range of pollutant kinds as well as one-, two-, and three-dimensional systems. In the US and other countries, it is one of the most extensively used models of water quality (Lin *et al.*, 2023). It may be linked to river basins and hydrodynamic models, enabling multi-year assessments under varied meteorological and environmental circumstances (Mohd *et al.*, 2022).

It has been implemented in Florida's key estuaries (Tuan, 2020). The Shenandoah River Basin in Virginia simulated nutrients, dissolved oxygen (DO), and chlorophyll dynamics using the Water Quality Analysis (WASP) simulation programme (Poursaeid, 2023). A sensitivity test's findings demonstrate that when model complexity and sensitivity increase, the model error reduces (Shamsuddin *et al.*, 2022). While total nitrogen and phosphorus tended to be overemphasised, the model predicted DO levels that were generally in line with the actual data (Lin *et*

al., 2023). The findings emphasise how temperature, flow, and speed affect water quality at different times of the year and different elevations in various river basin sections (Wu *et al.*, 2020).

2.2 AQUATOX

This is a simulation model for aquatic systems. AQUATOX predicts the fate of various pollutants, such as nutrients and organic chemicals, and their effects on the ecosystem, including fish, invertebrates, and aquatic plants (Nina & Galina, 2021) (Hameed *et al.*, 2016). The AQUATOX model has been used successfully to quantify the impacts induced by insecticides on ecosystem services provided by a lake from toxicity data for organism-level targets (Hanane *et al.*, 2022) (Fernandez *et al.*, 2022). The simulations showed how exposure to an insecticide could affect aquatic species populations (eg, abundance of recreational fish) and environmental properties (eg, water clarity) which in turn would affect the delivery of ecosystem services (Hinne *et al.*, 2020).

Using the AQUATOX EFDC model, a short-term environmental impact was assessed for scenarios of 30-30,000 kg of toluene runoff in South Korea's Jeonju River (Eddine *et al.*, 2023). The result of the AQUATOX-EFDC simulation showed a significant ecological impact, finding that in the scenario in which 3000 kg of toluene were leaked for one day, a substantial change was expected in the range 0-640 m from the accident site (Amanullah *et al.*, 2020). Besides, for a leak of 30,000 kg, a substantial change was expected in the range of 0-2300 m from the site of the accident, and the greatest damage was observed for the group of fish species, the apex predators (Ahmed *et al.*, 2019).

2.3 CE-QUAL-W2

This is a hydrodynamic, two-dimensional, longitudinal/vertical water quality model. The model works well for moderately long and narrow water bodies with longitudinal and vertical water quality variations because it assumes lateral homogeneity (Poursaeid & Poursaeed, 2023). Estuaries, lakes, reservoirs, and rivers have all been subjected to the model's application (Shin *et al.*, 2020). They used data from the Xiaksi River that was obtained in 2007 and 2008 to test the CE-QUAL-W2 model. The Xiaksi River's total mercury and methylmercury contents may be predicted using CE-QUAL-W2, according to the model's results (Hanane *et al.*, 2022). This application also showed how well the CE-QUAL-

W2 model could predict the intricate cycling and movement of mercury species in water systems (Yang & Moyer, 2020).

A hydrodynamic and water quality model of the Tigris River, CE-QUAL-W2, has been created because the total dissolved solids (TDS) in the river can exceed drinking water and irrigation lines by up to 1000 mg/L in Baghdad and other cities downstream. According to a sensitivity analysis, a 15% increase in river flow upstream causes the TDS content to drop by around 5% (Eddine *et al.*, 2023). It was advised to strictly regulate the flows through Lake Tharthar and the return flows of irrigation in the main trunk of the Tigris River to keep the total dissolved solids concentration below 500 mg/L (Kung & Mu, 2019). The average annual flow in the Tigris River in Baghdad was to be maintained above 420 m³/s.

The water quality in a reservoir of Uiam Lake in Korea, which gets nutrients from both non-point and point sources, was modelled using the two-dimensional model (CE-QUAL-W2). The outcome of the scenario simulations showed that a 62% reduction in the chlorophyll content in the lake would occur from lowering the P of the STP effluent from 0.9 mg/L to 0.1 mg/L (Eddine *et al.*, 2023). Depending on the hydrological parameters of the receiving water bodies and the runoff pattern, the impact of phosphorus loading on phytoplankton development can vary significantly (Than *et al.*, 2016).

2.4 Ansys CFX

This programme is the best in the market for calculating fluid dynamics in turbomachine applications. It is renowned for its exceptional speed, resilience, and precision (Nina & Galina, 2021). For applications like pumps, fans, compressors, and gas and hydraulic turbines, it is incredibly accurate and powerful while being incredibly easy to operate (Ni *et al.*, 2020). The propagation of turbidity currents was simulated using two alternative numerical codes with ANSYS CFX 17.1 and the Telemac 3D Solver (Poursaeid, 2023). Turbidity currents are recognised as a useful tool for controlling the buildup of fine sediments in tanks. Two setups were used to run the simulations.

The modelling of a turbidity current sinking is the subject of the first case. To control the tank's sedimentation, the second model involved evaluating the turbidity current ventilation. To give suggestions for modelling turbidity currents in actual tanks, the benefits and drawbacks of both techniques are examined (Abu *et al.*, 2023).

2.5 GWLF

This is a medium-level river basin loading model that was created to assess nonpoint source flow and the loading of nutrients and sediments from both urban and rural river basins (Nina & Galina, 2021). Using regions of different sizes (e.g., agricultural, wooded, and developed land), the model can simulate runoff, sediment, and nutrient loads (N and P) from a river basin. According to Morton and Henderson (2008), the model is continuous and uses daily time increments to calculate the water balance and meteorological data.

From 1997 to 2015, the Indian Mill Creek watershed in Michigan, USA, underwent simulations of water budgeting, field erosion, and stream erosion along a gradient of agricultural to urban land cover using the upgraded model of generalised river basin loading functions (GWLF-E). The outcomes demonstrated that the discharge into sub-pools may have been overstated by GWLF-E. GWLF-E may have trouble capturing the complexity of erosion rates, particularly in situations where sediment is deposited on current banks from upstream sources (Hanane *et al.*, 2022). In hydrographic basins, current bank erosion is sometimes underestimated by an order of magnitude, according to evaluation using current strip erosion data (Alsulaili & Refaie, 2020).

The RGWLF river basin model was used in the Luanhe river basin to look for spatial BMPs (Best Management Practices) for dissolved nitrogen (DisN). It is combined with a well-known multiobjective optimisation algorithm and the non-dominant genetic sorting algorithm II (NSGAI). The findings showed that the total dissolved nitrogen burden was 1.31×10^7 kg at a maximum cost of 1.77×10^8 yuan and 3.1×10^7 kg at a minimum cost of 9.09×10^7 yuan; in the absence of any measures, the DisN load was 4.05×10^7 kg (Bentéjac *et al.*, 2021).

2.6 WMS

This is a two-dimensional graphical system that simulates hydraulic and hydrographic basins (Poursaeid, 2023). It may be applied to the modelling of water quantity and quality. WMS was created at Brigham Young University in the United States in the early 1990s (Tiyasha *et al.*, 2021).

2.7 MONERIS

This model computes flow retention in the surface water network as well as emissions of phosphorus and nitrogen into surface water in various ways (Poursaeid

& Poursaeed, 2023). The Danube basin's nutrient inputs have been successfully modelled through the use of the MONERIS technique (modelling of nutrient emissions in river systems). Ighalo *et al.* (2020) describe it as a semi-static emission model that can be modified to handle priority compounds such as lindane and heavy metals (Poursaeid & Poursaeed, 2023). It applies to both point and diffuse nutrient sources. Nitrogen (N) and phosphorus (P) emissions in Poland's Oder and Vistula basins were calculated thanks to modelling studies (MONERIS), which also made it easier to estimate N and P retention in these basins between 1995 and 2015.

Population growth and agricultural intensification were the main causes of the 5,3 and 3,5-fold increases in N and P emissions in the Oder basin, with the highest levels (135,000 tonnes N year⁻¹ and 14,000 tonnes P year⁻¹) recorded at the start of 1980/1990. Throughout the economic transition era (which began in 1989), pro-ecological initiatives have been implemented in a variety of economic sectors, such as agriculture and environmental protection; for instance, by building a significant number of waste water treatment facilities (Hanane *et al.*, 2022). The Oder basin's emissions decreased from the aforementioned maximums to 94,000 tonnes N year⁻¹ and 5,000 tonnes P year⁻¹ between 1985 and 2015, while the Vistula basin saw a decrease in emissions from 170,000 to 140,000 tonnes N year⁻¹ and from 14,200 to 10,600 tonnes P year⁻¹ between 1995 and 2015 (Kouadri *et al.*, 2021).

2.8 QUAL2K

This is a modernised version of the QUAL2E (or Q2E) model (Brown and Barnwell 1987) for water quality in rivers and streams. Q2K and Q2E are comparable in the following ways: The channel is well-mixed laterally and vertically in a one-dimensional space (Osman *et al.*, 2021). Using the water quality model (Qual2k), simulations of the Dez River in Iran's water quality parameters (pH, NH₄-N, NO₃-N, dissolved oxygen, and electrical conductivity) were evaluated (Eddine *et al.*, 2023). The Qual2k model was calibrated using data collected in July 2012 and validated using data collected in March 2013.

For eight years, the water quality index was evaluated with consideration of drought and industry expansion. Iran's computed water quality index revealed values between 70 and 85. To sum up, the Qual2k model has been suggested as a suitable instrument for assessing and projecting future water quality (Kung & Wu, 2021). In Shixi Creek, southeast

China, a sparsely inhabited stream, a simulation-optimization technique based on the QUAL2K model has been created to offer nitrogen pollution mitigation solutions (Hanane *et al.*, 2022). With a normalised objective function (NOF) of less than 0,360, the model demonstrated strong agreement with field data from 22 sampling locations tested between March 2017 and February 2019 (Nina & Galina, 2021). Reducing more than 55% of N pollution from point sources and 10% of N pollution from nonpoint point sources is an optimal strategy to achieve the grade III water quality criteria for more than 80% of Shixi Creek's surface (Singha *et al.*, 2021).

2.9 SMS

Surface water modelling software is included in this bundle. US professionals from the Aquaveo Company created it. It can resolve both static and dynamic issues (Eddine *et al.*, 2023). Since it handles every step of the modelling process, from importing topographic and hydrodynamic data to visualising and analysing results, it is frequently used in process simulation in "river" type water systems (Lin *et al.*, 2023). River hydrodynamics, rural and urban floods, wave modelling, tracking the dynamics and physical characteristics of water particles, and pollution identification and analysis are all included in the modelling process (Chinh, 2019).

The Surface Water Modelling System (SMS) software's RMA2 module was used to conduct a numerical modelling study of current and tide to determine the tidal characteristics and currents in Lemong waters, Lampung province, as well as the parameters for coastal protection design (Nina & Galina, 2021). Using four levels of network resolution, the modelling was carried out using the online nesting approach, with the maximum resolution at 1 x 1 km² close to the Lemong beach waters. According to the modelling findings, the predominant mixed semi-diurnal type of tide in the Lemong Beach region reaches 1.4 m (Hanane *et al.*, 2022). Additionally, according to Ibrahim *et al.* (2023), the model indicates that the current pattern in the seas around Lemong Beach primarily flows northwest during ebbs and southeastward during flows.

2.10 WQRSS

The United States Army Corps of Engineers created this model, which replicates a variety of aquatic biota, including plankton, algae, bacteria, coliforms, and various fish species, as well as DO, total dissolved solids, P, NH₃, NO₂⁻, and NO₃⁻, alkalinity, and total

carbon (Lin *et al.*, 2023). It establishes depths and speeds as well as the hydrodynamic form (Chinh, 2019). A reservoir version of the Ecological Water Quality for River Systems (WQRRS) model was used to examine the water quality of Lake LaFarge, Wisconsin. Temperature, dissolved oxygen, and algae densities were the main topics of the investigation.

The results of the studies showed that phosphorus was probably the limiting nutrient in the suggested network and that all of the soluble phosphorus and nitrogen in the river were essentially accessible for algal development (Eddine *et al.*, 2023). According to the calculations, Lake LaFarge is likely to experience thermal stratification from May to September (Lap *et al.*, 2023). Various techniques are available for simulating the movement of contaminants in rivers. These comprise conceptual tank-type models and full hydrodynamic models built in InfoWorks-RS (Innovyze), HEC-RAS (USA), and MIKE11 (DHI Water & Environment) (Ahmad *et al.*, 2021).

2.11 MIKE 11

For modelling flows, water quality, and sediment movement in estuaries, rivers, irrigation systems, canals, and other water bodies, this software package is designed for engineers (Hanane *et al.*, 2022). For thorough study, design, administration, and operation of both basic and large river and canal systems, this user-friendly, one-dimensional, fully dynamic modelling tool is available (Mokhtar *et al.*, 2022). A leak simulation model was created using the MIKE 11 NAM conceptual hydrological technique for the Arpasub Basin in the Seonath River Basin in Chhattisgarh, India.

The model's ability to describe the precipitation drainage process from the basin and, consequently, forecast the daily flow, is demonstrated by the calibration and validation findings (Khoi *et al.*, 2019). The one-dimensional hydraulic modelling tool MIKE11 was used to analyse the sensitivity of the Manning roughness coefficient on the water level of the Delta Mahanadi rivers (Nina & Galina, 2021). The model's sensitivity to the roughness coefficient was demonstrated by the findings. As the stage grows, the value "n" rises (Ye *et al.*, 2019).

Info Works ICM (Integrated Catchment Modelling): According to Khoi *et al.* (2022), this software package is the first of its kind on the market for fully integrated 1D/2D hydrodynamic modelling of sewage systems and rivers. Hydrographs were simulated at the outflow of a basin in Shenzhen, China (drainage

area 8.3 ha, with 95% permeable regions) using the InfoWorks Integrated Catchment Modelling (ICM) model (Poursaeid, 2023). It has been demonstrated that, for precipitation events with deeper depths and longer durations, the model integrating the direction of the quasi-linear double reservoir has a bigger influence on simulated hydrographs than the US EPA's nonlinear reservoir steering technique (Huong, 2018). By utilising the InfoWorks ICM model, a novel design approach was suggested for the optimal placement of water-quality storage tanks (Poursaeid & Poursaeid, 2023).

Accordingly, the suggested approach may lower the overall volume of decentralised storage tanks to 0.38 times that of a terminal tank when applied to the construction of decentralised storage tank locations in Fuzhou, China (Abba *et al.*, 2020). Because of the harm that floods inflict on natural resources, it is now essential to map and monitor floods (Eddine *et al.*, 2023). This phenomenon is common in L'Île-Bouchard, and residents of the Vienne's flood-prone districts are particularly exposed to flooding. Using ArcMap and InfoWorks ICM software, a study of the flood hazard and risk maps of a river sector for the Vienne River, which passes through this commune, was conducted (Lin *et al.*, 2023). According to the findings, the model can replicate the depth of flooding that occurs in real life, which allows for the development of flood risk and hazard maps (Aldhyani *et al.*, 2020).

2.12 HEC-RAS

With the help of this programme, one can do continuous one-dimensional flow, one- and two-dimensional non-uniform flow calculations, sediment transport and mobile bed calculations, and modelling of water temperature and quality (Sulaiman *et al.*, 2018). To simulate the carbonaceous biochemical oxygen demand (CBOD) and dissolved oxygen (DO) in a chosen region that stretched to about 25 km in the Diyala River and 22 km in the river Tiger, a numerical model of water quality (HEC-RAS) was created (Hanane *et al.*, 2022). According to the findings, the Diyala River was contaminated in the area downstream of the Al-Rustimiya wastewater treatment facilities, with CBOD levels varying between 18 and 25 mg/L and DO values between 1 and 3,1 mg/L during the wet and dry seasons.

In contrast, CBOD levels in the Tigris River dropped from 3 to 5 mg/L during the rainy season and from 17 to 18,1 mg/L during the dry season (Ahmed *et*

al., 2021). The cleaning of a bridge over the Kabul River in the vicinity of Peshawar, Pakistan, was examined using the HEC-RAS model. The findings demonstrated that the cleaning depth produced by the square pillars was higher than that of the circular pillars. Regarding hole diameters, a similar pattern was noted (Aminu, 2022). The hydraulic software (HEC-RAS) and the experimental results were compared, and it was shown that under comparable circumstances, the hydraulic software produced somewhat greater cleaning depth values (Nina & Galina, 2021). However, when discharge increased, this disparity shrank (Schapire, 2003).

2.13 TLM

To make quick and efficient decisions in an emergency, this transmission line matrix model may be used to forecast the spatiotemporal evolution of a contaminant in natural flows (Joseph, 2022). In a nutshell, the digital techniques for reducing digital water pollution were studied. Based on a comparison of software tools according to the number of model parameters, we were able to determine that software tools with more parameter modelling are more complex that is, they have more features. The complexity of the aquatic ecosystems under research was another crucial factor in the analysis process, necessitating the use of several software tools for the simulations of each aquatic environment. Both geographical and temporal information are crucial when assessing water contamination using digital approaches.

It has been determined that not every product examined can be used in every situation. Both the pre-design phase and the actual usage process of aquatic systems make use of mathematical and numerical models. Digital methods are effective tools for assessing water quality and creating scenarios that might foretell water contamination.

2.14 Data Mining Approaches to Water Pollution

In the previous three years, from 2019 to 2021, some machine learning algorithms were employed for water quality monitoring systems. The author claims that the most popular machine learning (ML) modelling technique used for tracking water quality is artificial neural network (ANN) modelling (Ighalo *et al.*, 2020). With a limited quantity of data, ANN approaches may produce excellent prediction results and offer higher resilience and calibration that is more accessible for processing complicated and nonlinear datasets (Ighalo *et al.*, 2020).

There are several applications for applying machine learning algorithms to predict surface water quality worldwide. A ten-year evaluation of the literature on water quality indices in the field of artificial intelligence (AI) was conducted to identify the most practical or suitable models and techniques for use by researchers. Although artificial intelligence (AI) has been used much more in the field of water quality in the past ten years (Aminu, 2022), there is still an opportunity for researchers to become engaged and enhance the calculations, forecasts, and other aspects of the water quality index.

The irrigation water quality index of the Bahr El-Baqr region has been forecasted by certain case studies, such as the one based on the machine learning and regression model of Mokhtar *et al.* (2022). According to Egypt's study findings, the stepwise regression model is the most effective prediction model, followed by partial least squares regression (PLS), principal component regression (PCR), and others (Egypt). Babbar & Babbar (2017) used data mining approaches (k-nearest neighbour, decision tree, naive Bayes, artificial neural network, support vector machine) for the prediction of river water quality index. The findings indicate that the best prediction models are thought to be decision trees and support vector machines.

Yadav *et al.* (2021) computed the irrigation water quality index (IWQI) using five water quality indicators in their IoT-based water quality index prediction for farm irrigation. Three parameters were obtained from a set of five using the correlation analysis approach. The best classification model for forecasting water quality, according to the results, is the random forest model. The study employed five machine learning models to predict irrigation water quality indicators, with the SVM model being the most appropriate for all irrigation indicators. Prediction of irrigation water quality indices based on machine learning algorithms has also been applied in semiarid environments (Dimple *et al.* 2022).

Additionally, there has been an improvement in the prediction of the water quality index. Mohd *et al.* (2022) employed eight machine-learning regression models based on historical data from Indian river systems to forecast the index. The outcomes demonstrate that the best performance is provided by the ridge and linear regression models. Bui *et al.* (2020) found that a new hybrid machine learning algorithm improved the prediction of water quality indices by using 12 hybrid and 4 separate machine

learning algorithms to predict solely water quality indicators. Iran's surface water quality.

The results show that the best input matching models and the BA-RT matching approach outperform the others. In the El Kharga Oasis in Egypt's Western Desert, some irrigation water quality indices (IWQIs) and geographic information systems (GIS) were utilised to evaluate the groundwater (GW) quality for agricultural land. Ibrahim *et al.* (2023) forecasted groundwater quality for irrigation using GIS techniques, machine learning models, and integrated water quality metrics. Two machine learning (ML) models, the adaptive neuro-fuzzy inference system (ANFIS) and the support vector machine (SVM) were used to predict eight IWQIs.

Based on prediction skill criteria, the simulation models' performance was assessed, and it was found that the ANFIS and SVM models could reasonably accurately simulate the IWQIs. Support vector machines (SVMs) linked with water quality indices (WQI) were used by Abu *et al.* (2023) to merge a machine learning-based model with WQI for groundwater quality evaluation. In comparison to the SVM model and WQI, the SVM-WQI model displays a low proportion of the region for outstanding class. Overall, the understanding of water quality assessment provided by the combined ML model and WQI may prove useful in the future development of these places.

The study assesses the effectiveness of machine learning models for multiclass classification in water quality assessment and evaluation and finds that SVM is the best model to predict river water quality. This is in addition to the classification of the water quality index based on a machine learning model for the Langat River basin (Shamsuddin *et al.*, 2022). supervised machine learning models were used by Fernandez del Castillo *et al.* (2022) to forecast the ecosystem water quality index and classify the water quality of a badly contaminated river (supervised machine learning).

By reducing the number of water quality measures from 17 to 12, the water quality monitoring programme may be expanded while still predicting the ecosystem's water quality index (SGR-WQI). Santiago-Guadalajara River's current water level (Mexico). Water quality indexes have also been predicted and sorted using deep learning algorithms. In 2021, Tiyasha *et al.* used an artificial intelligence model to forecast the river water quality index. The results indicated that, for both large-scale and small-

scale watershed datasets, the H₂O deep learning model was the most accurate, followed by a random forest model. Hameed *et al.* (2016) predicted the water quality index using artificial intelligence algorithms. Accurate prediction of the water quality index (WQI) is possible using an ANN.

In tropical areas (like Malaysia), the radial basis functional neural network (RBFNN) model is thought to be the most accurate at forecasting WQI. Given how time-consuming manual calculation approaches are, the suggested method offers an efficient substitute for calculating and estimating the WQI. An artificial intelligence (AI) programme was created by Aldhyani *et al.* (2020) to forecast the water quality index (WQI) and water quality classification (WQC). The findings demonstrate that the suggested models can correctly categorise water quality and forecast the WQI. SVM, K-NN, and Naive Bayes are a few machine learning algorithms and artificial neural network models that may be used to effectively forecast the water quality classification (WQC) and the water quality index (WQI).

When it came to WQI value prediction, the NARNET model outperformed the LSTM by a small margin, while the SVM method produced the best results (97.01%) for WQC prediction. Ahmed *et al.* (2019) estimated the water quality index (WQI) using a supervised machine learning technique as well. The most effective algorithms, according to the results, are polynomial regression and gradient enhancement (MAE values of 2.7273 and 1.9642, respectively). The most effective method for classifying water quality grades is the multilayer perceptron (MLP) (WQC). The suggested approach is appropriate for real-time water quality detection systems as it delivers a respectable level of accuracy while using the fewest possible parameters.

The La Buong River in Vietnam has been using machine learning algorithms to forecast the water quality index (Khoi *et al.* 2022). The efficacy of twelve machine learning models in forecasting the water quality index is assessed in this study. All 12 models perform well in predicting the WQI, according to the data, however the XGBoost model has the greatest accuracy ($R^2= 0.989$ and $RMSE= 0.107$). Tan *et al.* (2016) estimated the water quality index of the Dong Nai River, which flows through the provinces of Dong Nai and Binh Duong, using an artificial neural network (ANN). The study's findings show that when compared to the actual value of the water quality index (WQI), the projected WQI is highly significant

and has a high correlation coefficient ($R= 0.974$ and $p= 0.000$). Additionally, ANN models outperform multivariate regression methods in terms of predictive values.

To summarise, prior research on the use of deep learning in water quality forecasting has mostly concentrated on estimating the water quality index (WQI) and forecasting water quality metrics, primarily physical factors. Additionally, studies have shown highly promising outcomes when combining real-time monitoring networks with deep learning algorithms. Nevertheless, no study has used the technique of picking crucial variables from hundreds of water quality metrics (monitoring) as input data to determine the machine learning and deep learning models' surface water quality index (WQI). Furthermore, this research is well-known around the globe. The potential of deep learning and machine learning algorithms in anticipating the surface water quality index (WQI) based on data input (minimum water quality parameter) to lower the cost of surface water quality monitoring is not well-studied in poor nations.

2.15 Comparison of the two Approaches

Because machine learning and deep learning algorithms can handle vast quantities of data and produce very precise predictions, they have been used more and more recently across the globe to calculate and forecast water quality indices. Algorithms for machine learning and deep learning can effectively handle multidimensional and missing data, as well as nonlinear connections between water quality measures. Furthermore, when fresh data becomes available, these algorithms may constantly improve their predictions by learning from the data in real time.

BOD may be predicted and wastewater treatment plant performance processes can be identified using ANN modelling settings (Alsulaili & Refaie, 2020). Prediction models can facilitate online control systems and save time when estimating BOD. Since it is less complicated than other ML algorithms, other ML approaches like multiple linear regression (MLR) and random forest (RF) are typically employed as the methodology. Hybrid models,

which are created from many traditional ML models and combined with optimisation techniques, represent the new wave of ML. It is used to enhance performance and provide accuracy, calculation, and functionality from a single model (Ardabili *et al.*, 2019). It has

been demonstrated that this approach offers several exceptional benefits over traditional methods.

To build prediction models, data-driven virtual sensing techniques require several inputs and readily quantifiable characteristics. It is possible to measure output parameters like total phosphorus (TP), sodium absorption ratio (SAR), total nitrogen (TN), magnesium absorption ratio (MAR), and residual sodium carbonate (RSC) using input factors like pH, electrical conductivity (EC), temperature, turbidity, and DO. For instance, we require COD as one of the inputs to estimate the TP and TN. The water quality characteristics can be converted by the sensor module submerged in the water into a corresponding, quantifiable electrical amount that is sent to the coordinating module. The usage of trustworthy and accurate sensors is crucial as it influences efficiency. Put differently, due to a variety of factors such as process nonlinearity, inadequate sample sizes, improper input parameter sensor selection, and others, the predicted accuracy of virtual sensing rapidly declined.

2.16 Social Media Data Mining and Water Pollution

Le-Coz *et al.* (2020) and Starkey *et al.* (2019) are among the few studies that have explored the application of this kind of data for hydrological study and modelling. Even less research has looked at the creation of monitoring records for ungauged water bodies using data from social media (Michelsen *et al.*, 2019). Below are some examples of these studies: Barker & Macleod (2019) used real-time river levels across the United Kingdom to monitor flood occurrences and water pollution through an analysis of Twitter data. A logistic regression-based classifier and paragraph vectors were employed in the creation of a Twitter data mining pipeline. To provide stakeholders with a more comprehensive picture of the local situation, the collected flood and water-polluted data were combined with real-time environmental data.

In a different study, Bischke *et al.* (2019) adopted and enhanced the data by utilising photos from Twitter to gauge the intensity of floods. They utilised satellite photographs to detect water pollution and floods. Firstly, high-resolution satellite images were also examined for high flood levels and high particles on the water indicating water pollution levels by automatically detecting water levels, flowing particles, and generating a flood map. After pre-processing Twitter images by removing duplicate images, an

algorithm was created to identify those images and the water level as a result of flood and pollution.

Albuquerque *et al.* (2020) used statistical analysis to identify spatial patterns in the flood and pollution-related tweets and combined it with authoritative data by analysing a case study of the River Elbe Flood in Germany in June 2013. This allowed them to determine the relevancy of georeferenced social media messages from Twitter during any flood and pollution event. The findings indicated that there is a greater likelihood of a specific flood incidence and water particles being associated with tweets sent within a 10-kilometre geographical region. In Jakarta, Indonesia, the “wisdom of the crowd” method was compared to the observed watershed activity using the number of Tweets that revealed consistent patterns in the data (Eilander *et al.*, 2019).

This strategy works better in places where there is a higher concentration of social media users. Following the creation of flood and water pollution maps, the observed data on the georeferenced activity of Twitter users in a given area was overlaid on a Digital Elevation Model (DEM) together with pollution level and flood depth measurements. There was clear and consistent evidence of water pollution in that location from the tweets containing the depth of the water.

Karmegam *et al.* (2021) carried out a similar study in which they mapped flood depth using data from social media, and the results were verified using real-time flow data. Le-Coz *et al.* (2020) carried out a more thorough investigation in Argentina, France, and New Zealand, where they created a special Flood Chasers Project website. Uploading pictures and videos of any floods or polluted water was encouraged. Large Scale Particle Image Velocimetry (LSPIV), an efficient method for post-flood discharge and pollution estimation, was used to model river flow velocity and discharge from these images using PIV/PTV analysis tools. This allowed for the mapping and estimation of floods and polluted water.

By using these methods, citizen science has the potential to be used for assessing pollution and flood risks. Additional research utilising VGI on water pollution and flood damage has been documented in the literature. These studies make use of specialised platforms like PetaJakarta in Jakarta, the QLF flood crisis map in Australia (Koswate *et al.*, 2019), and flooding locations in Brazil (Hirata *et al.*, 2020). Using data from Twitter and reliable sources, Restrepo-Estrada *et al.* (2019) investigated the use

of social media for rainfall-runoff calculations and forecasting floods and their cleansed condition.

Achieving 71% accuracy required combining real-time official flood measurements with geolocation Twitter data as an input for the Probability Distribution Model (PDM), which has nearly doubled in precision with the addition of social media data. Rosser & Leibovici (2019) used a Bayesian statistical model to analyse Flickr posting activity, remote sensing, and topographic map data from the 2014 UK flood to create a probability map that shows the likelihood of floodwater and pollutants present. The map predicts floods based on seven parameters, including rainfall, area, pressure, velocity, gauge, average temperature, and average wind speed.

Deep neural networks were utilised by Panigrahi *et al.* (2019). Deep learning models were used to study the Daya and Bhargavi rivers in India to anticipate the discharge volume. According to the findings, the Local Linear Radial Basis Functional Neural Network (LLBRFNN) performed well in predicting pollution and flood levels and had the lowest Mean Absolute Percentage Error (MAPE) and Mean Square error (MSE). In a separate research, Panigrahi *et al.* (2018) used the same rivers to build the Cascaded Functional Link Artificial Neural Network (C-FLANN) and update the model’s parameters using Harmony Search (HS) and Differential Evolution (DE). The results showed that c-FLANN trained with HS could predict the water level and purification state more correctly, using the same seven atmospheric parameters that were used to predict river water flow.

The categorization of massive volumes of data in a variety of formats (text, video, and picture) and the extraction of pertinent metrics from the data are the main tasks involved in employing social media for water pollution monitoring. To help with this endeavour, machine-learning techniques are typically employed. For instance, Huang *et al.* (2022) used supervised logistical regression with an unsupervised machine-learning technique called clustering to group emergency-related material according to similarity and group postings according to various occurrences. According to Gallego *et al.* (2019), ResNet is a pre-trained model for image categorization. Feature extraction, picture classification, image segmentation, and object identification are common applications for the ResNet architecture. Deep convolutional neural networks (CNNs) are used in this technique for both identification and classification.

A machine-learning approach called random forest is utilised for regression and classification (Adeniran, Folorunso et al., 2024). It is a technique that builds many decision trees. Individual decision trees are created during the training phase by choosing at random the characteristics at each node that establish the split. The individual weights assigned to each tree during categorization are then given. According to Nair et al. (2019), random forests are capable of handling missing values and data including outliers. According to Nair et al. (2019), the Naïve Bayes statistical classifier offers conditional independence between predictors. Its naivety stems from the fact that it is primarily predicated on the notion that all predictors or traits are conditionally independent (Aharwal, 2019).

A support vector machine (SVM) classifier is implemented using sequential minimum optimisation (SMO) on the Weka (Waikato Environment for Knowledge Analysis) platform. By creating an N-dimensional hyperplane that can effectively divide data into two groups, it is designed for numerical prediction and data classification (Ayodele, 2021). Because SVM may eliminate the necessity for feature selection, it performs well in text classification problems (Alshutayri et al., 2019). Based on information theory, the classification algorithm C4.5 generates a decision tree. It induces decision trees for classification using a greedy strategy and the information entropy idea (Adeniran, Adeniran et al., 2024; Adeniran, Onuajah et al., 2024). It utilises information entropy to build a decision tree from labelled training data and allows nominal classes (Aharwal, 2019).

3. Conclusion

Previous research has concentrated on creating a framework for social media activity in water pollution, especially in places that flood. A few of these studies also employed volunteer geographic data to examine these types of occurrences at the hydrological catchment scale. The usefulness of such data for pollution forecasts and the data quality of information gathered from social media platforms must be addressed, especially for desert regions with intermittent but heavy rainfall and a significant risk of polluted areas.

This study examines reports and records of water pollution from a variety of social media sites, including YouTube, Flickr, Facebook, Instagram, and Twitter, for the first time. It verified data on water pollution from social media. Based on the results of

many machine learning classifiers, social media data was validated to predict and evaluate water pollution. The area under the curve (AUC), kappa statistics, and root-mean-square error (RMSE) will be utilised for this purpose. We will analyse the quality of the data from Twitter, Facebook, Instagram, YouTube, and Flickr, and use kappa statistics, AUC, and RMSE to evaluate the model's performance.

4. References

1. Abba, S. I., Pham, Q. B., Saini, G., Linh, N. T. N., Ahmed, A. N., Mohajane, M., Khaledian, M., Abdulkadir, R. A. & Bach, Q. V. (2020). Implementation of data intelligence models coupled with ensemble machine learning for prediction of water quality index. *Environment Science Pollution Research*, 27: 41524-41539.
2. Abu, S. A. , Ismael, I. S., El-Sabri, M. A. S., Abdo, M. S. & Farhat, H. I. (2023). Integrated machine learning-based model and WQI for groundwater quality assessment: ML, geospatial, and hydroindex approaches. *Environment Science Pollution Research*, 30(18):53862-53875. <https://doi.org/10.1007/s11356-023-25938-1>
3. Adeniran, A. O., Folorunso, S. S., Familusi, O. B., & Mbachu, O. V. (2024). Adoption of Fourth Industrial Revolution Technologies in the Construction Sector: Evidence from Previous Studies. *International Journal of Innovation in Engineering*, 4(3), 1-18. <https://doi.org/10.59615/ijie.4.3.1>
4. Adeniran, A. O., Adeniran, A. A., & Udorah, F. N. (2025). Nexus between Economic Development and Renewable Energy on CO2. *Management Analytics and Social Insights*, 2(1), 1-11. <https://doi.org/10.22105/masi.v2i1.55>
5. Adeniran, A. A, Adeniran, A. O, Familusi, O. B., & Adedayo, O. (2024). The Outlook of Cyber Security in African Businesses: Issues and Way-out. *Management Analytics and Social Insights*, 1 (2), 260-271.
6. Adeniran, A. O., Onuajah, S. I., Adeniran, A. A., and Ogunmola, M. A. (2024). Implementing cloud centric IOT transformations: merits and demerits. *Systemic Analytics*, 2 (2), 174-187.
7. Adeniran, A. O., Oyeniran, G. T., Adeniran, A. A., & Mosunmola, M. J. (2024). Digitization in Logistics and its Effect on Sustainability in Nigeria. *Discovery 2024*; 60: e15d1420
8. Adeniran, A. O., Sidiq, O. B., Adeniran, A. A., & Oyeniran, G. B. (2024). Sustainability Impact of Digital Transformation in E-Commerce Logistics. *International Journal of Innovation in Marketing Elements*, 4(1), 1-18.

9. Aharwal, R. P. (2019). Evaluation of various classification techniques of weka using different datasets. *International Journal of Advanced Research, Innovation, Ideas & Education*, 2;4396: 2395.
10. Ahmad, W., Iqbal, J., Nasir, M. J., Ahmad, B., Khan, M. T., Khan, S. N. & Adnan, S. (2021). Impact of land use/land cover changes on water quality and human health in district Peshawar Pakistan. *Science Rep*, 11(1):16526. <https://doi.org/10.1038/s41598-021-96075-3>.
11. Ahmed, M., Mumtaz, R. & Hassan, S. M. (2021). Analysis of water quality indices and machine learning techniques for rating water pollution: a case study of Rawal Dam, Pakistan. *Water Supply*, 21 (6): 3225–3250. <https://doi.org/10.2166/ws.2021.082>.
12. Ahmed, U., Mumtaz, R., Anwar, H., Shah, A. A., Irfan, R. & Garcia-Nieto, J. E. (2019). Efficient water quality prediction using supervised machine learning. *Water*, 11(11):2210. <https://doi.org/10.3390/w11112210>
13. Akinloye, D. N. (2024a). Grandparent-grandchild Communication Synthesis of Integrative Model. *Psychology Nexus*, 1(1), 1-15.
14. Akinloye, D. N. (2024b). Communication in Divorced Families with Adult Children: A Literature. *Annals of Journalism and Mass Communication*, 3(1), 70-79.
15. Akinloye, D. N. (2025). Communication Patterns and the Influence of Social Media Contents on Young Adults' Wellbeing in Texas, United States of America. *Journal of Communication*, 6(2), 1-22.
16. Albuquerque, J. P. de-Herfort, B., Brenning, A., & Zipf, A. (2020). A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographic Information Science*, 29: 667-689. <https://doi.org/10.1080/13658816.2020.996567>.
17. Alshutayri, A., Atwell, E. S., Alosaimy, A., Dickins, J., Ingleby, M., & Watson, J. (2019). Arabic language WEKA-based dialect classifier for Arabic automatic speech recognition transcripts. Available: <http://eprints.whiterose.ac.uk/107396/>. Proc Third Workshop NLP Similar Lang Var Dialects VarDial 2016. Leeds; pp. 204-211.
18. Aldhyani, T. H. H., Al-Yaari, M., Alkahtani, H. & Maashi, M. (2020). Water quality prediction using artificial intelligence algorithms. *Applied Bionics Biomech*, 1-12. <https://doi.org/10.1155/2020/6659314>
19. Alsulaili, A. & Refaie, A. (2020). Artificial neural network modelling approach for the prediction of five-day biological oxygen demand and wastewater treatment plant performance. *Water Supply*, 21, 1861-1877.
20. Amanullah, Khalid, S., Imran, Khan, H. A., Arif, M., Altawaha, A. R. & Parmar, B. (2020). Effects of Climate Change on Irrigation Water Quality. In S. Fahad, M. Hasanuzzaman, M. Alam, H. Ullah, M. Saeed, I. Ali Khan, & M. Adnan (Eds.), *Environment, Climate, Plant and Vegetation Growth* (pp. 123-132).
21. Aminu, I. I. (2022). A novel approach to predict water quality index using machine learning models: a review of the methods employed and future possibilities. *Global Journal of Engineering Technology & Advancement*, 13(2):026-037. <https://doi.org/10.30574/gjeta.2022.13.2.0184>.
22. Ardabili, S., Mosavi, A. & Várkonyi-Kóczy, A. R. (2019). Advances in Machine Learning Modeling Reviewing Hybrid and Ensemble Methods. In *Engineering for Sustainable Future. Inter-Academia; Lecture Notes in Networks and Systems; Springer: Cham, Switzerland, 2022; Volume 101*.
23. Ayodele, T. O. (2021). Types of machine learning algorithms. *New Advanced Machine Learning*, IntechOpen. <https://doi.org/10.5772/9385>.
24. Babbar, R. & Babbar, S. (2017). Predicting river water quality index using data mining techniques. *Environment & Earth Science*, 1-12. <https://doi.org/10.1007/s12665-017-6845-9>.
25. Barker, J. L. P., & Macleod, C. J. A. (2019). Development of a national-scale real-time Twitter data mining pipeline for social geodata on the potential impacts of flooding on communities. *Environment Modelling Software*, 115: 213-227. <https://doi.org/10.1016/j.envsoft.2019.11.013>.
26. Bentéjac, C., Csorgo, A. & Martínez-Munoz, G. (2021). A comparative analysis of gradient boosting algorithms. *Artificial Intelligence Review*, 54(3):1937-1967. <https://doi.org/10.1007/s10462-020-09896-5>.
27. Bischke, B., Helber, P., Zhao, Z., De-Bruijn, J., & Borth, D. (2019). The multimedia satellite task at MediaEval 2018: Emergency response for flooding events. *Work Notes. Proc Mediaev Workshop Mediaev*. Available at: <https://www.alexandria.unisg.ch/id/eprint/258247>. pp. 1–3.
28. Bui, D. T., Khosravi, K., Tiefenbacher, J., Nguyen, H. & Kazakis, N. (2020) Improving prediction of water quality indices using novel hybrid machine-learning algorithms. *Science Total Environment*, 721:137612. <https://doi.org/10.1016/j.scitotenv.2020.137612>.
29. Chinh, V. Q. (2019). Monitoring and forecasting water quality in the irrigation system of Bac Duong for agricultural production in 2019. Retrieved from Ha-Noi, D., Rajput, J., Al-Ansari, N. & Elbeltagi, A. (2022). Predicting irrigation water quality indices based on data-driven algorithms: a case study in a

- semiarid environment. *Journal of Chemistry*, 2022:1-17. <https://doi.org/10.1155/2022/4488446>.
30. Eddine, Y. T., Marouf, N., Sehtal, S. & Jaafari, A. (2023). Integrating Support Vector Machines with Different Ensemble Learners for Improving Streamflow Simulation in an Ungauged Watershed. *Water Resources Management*, 38(11):1-15. Doi:10.1007/s11269-023-03684-w.
 31. Eilander, D., Trambauer, P., Wagemaker, J., & van-Loenen, A. (2019). Harvesting social media for generation of near real-time flood maps. *Procedia Engineering*, 154: 176-183. <https://doi.org/10.1016/j.proeng.2019.07>.
 32. Familusi, O. B., Omoyeni, O. D., Samchuks, O. M., & Adeniran, A. O. (2024). Effect of Renewable Energy on CO₂ Emission in Sub Saharan Africa. *Systemic Analytics*, 2 (2), 304-314.
 33. Fernandez, A., Yebra-Montes, C., Verduzco, M., Anda, J. E., Garcia-Gonzalez, A. & Gradilla-Hernández, M. S. A. (2022). Simple prediction of an ecosystem-specific water quality index and the water quality classification of a highly polluted river through supervised machine learning. *Water*, 14(8):1235. <https://doi.org/10.3390/w14081235>.
 34. Gallego, A. J., Pertusa, A., & Gil, P. (2019). Automatic ship classification from optical aerial images with convolutional neural networks. *Remote Sensing*, 10: 511. <https://doi.org/10.3390/rs10040511>.
 35. Hameed, M., Sharqi, S. S., Yaseen, Z. M., Afan, H. A., Hussain, A. & Elshafe, A. (2016). Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in the tropical region, Malaysia. *Neural Computational Application*, 28(S1):893- 905. <https://doi.org/10.1007/s00521-016-2404-7>.
 36. Hanane, R., Karima, K. & Khalid E. (2022). Water Quality Modeling in Atlantic Region: Review, Science Mapping and Future Research Directions. *Water Resources Management*, z37(1). Doi:10.1007/s11269-022-03382-z.
 37. Hirata, E., Giannotti, M. A., Larocca, A. P. C., & Quintanilha, J. A. (2020). Flooding and inundation collaborative mapping use of the Crowdmap/Ushahidi platform in the city of Sao Paulo, Brazil. *Journal of Flood Risk Management*, 11: S98–S109. <https://doi.org/10.1111/jfr3.12181>.
 38. Huang, L., Shi, P., Zhu, H., & Chen, T. (2022). Early detection of emergency events from social media: A new text clustering approach. *Nat Hazards*, 1-25. <https://doi.org/10.1007/s11069-021-05081-1z>.
 39. Huong, V. T. T. (2018). Research and propose solutions to reduce water pollution in the Bac Hung Hai irrigation system. Retrieved from Ha Noi Hussain D, Khan AA (2020) Machine learning techniques for monthly river flow forecasting of Hunza River, Pakistan. *Earth & Scientific Information*, 13(3):939-949. <https://doi.org/10.1007/s12145-020-00450-z>.
 40. Ibrahim, H., Yaseen, Z. M., Scholz, M., Ali, M., Gad, M., Elsayed, S., Khadr, M., Hussein, H., Ibrahim, H. H., Eid, M. H., Kovács, A., Péter, S. & Khalifa, M. M. (2023). Evaluation and prediction of groundwater quality for irrigation using integrated water quality indices, machine learning models and GIS approaches a representative case study. *Water*, 15(4):694. <https://doi.org/10.3390/w15040694>.
 41. Ighalo, J. O., Adeniyi, A. G. & Marques, G. (2020). Artificial intelligence for surface water quality monitoring and assessment: A systematic literature analysis. *Modeling Earth System and Environment*, 7, 669-681.
 42. Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis on Data Mining, ASA and Data Scientific Journal*, 15(4):531–538. <https://doi.org/10.1002/sam.11583>.
 43. Khoi, D. N., Nguyen, V. T., Sam, T. T. & Nhi, P. T. (2019). Evaluation of effects of climate and land-use changes on streamflow and water quality in the La Buong River Basin, Southern Vietnam. *Sustainability*. <https://doi.org/10.3390/su11247221>.
 44. Khoi, D. N., Quan, N. T., Linh, D. Q., Nhi, P. T. T. & Thuy, N. T. D. (2022). Using machine learning models for predicting the water quality index in the La Buong River, Vietnam. *Water*, 14(10):1552. <https://doi.org/10.3390/w14101552>.
 45. Karmegam, D., Ramamoorthy, S., & Mappillairaju, B. (2021). Near real-time flood inundation mapping using social media data as an information source: a case study of 2015 Chennai flood. *Geoenvironmental Disasters*, 8:1–11. <https://doi.org/10.1186/s40677-021-00195-x>.
 46. Koswatte, S., McDougall, K., & Liu, X. (2019). XSDI and crowdsourced spatial information management automation for disaster management. *Survey Review*, 47: 307-315. <https://doi.org/10.1179/1752270615Y>.
 47. Kouadri, S., Elbeltagi, A., Islam, A. R. M. T. & Kateb, S. (2021). Performance of machine learning methods in predicting water quality index based on irregular dataset: application on Illizi region (Algerian southeast). *Applied Water Science*, 11(12):190. <https://doi.org/10.1007/s13201-021-01528-9>.
 48. Kung, C. C. & Mu, J. E. (2019). The prospect of China's renewable energy development from pyrolysis and biochar applications under climate change. *Renewable & Sustainable Energy Review*., 114:109343. <https://doi.org/10.1016/j.rser.2019.109343>.

49. Kung, C. C. & Wu, T. (2021). Influence of water allocation on bioenergy production under climate change: a stochastic mathematical programming approach. *Energy*, 231:120955. <https://doi.org/10.1016/j.energy.2021.120955>.
50. Lap, B. Q., Phan, T. T. H., Nguyen, H. D., Quang, L. X., Hang, P. T., Phi, N. Q., Hoang, V. T., Linh, P. G. & Hang, B. T. T. (2023). Predicting Water Quality Index (WQI) by feature selection and machine learning: a case study of An Kim Hai irrigation system. *Ecol Inform* 74:101991. <https://doi.org/10.1016/j.ecoinf.2023.101991>
51. Le-Coz, J., Patalano, A., Collins, D., Guille, N. F., Garcı, C. M., & Smart, G. M. (2020). Crowdsourced data for flood hydrology: Feedback from recent citizen science projects in Argentina, France and New Zealand. *Journal of Hydrology*, 541: 766-777. <https://doi.org/10.1016/j.jhydrol.2020.07.036>.
52. Lin, C. H. & Chen, W. H. (2023). A Technology-Organization-Environment (TOE) Framework Based on Scientometry for Understanding The Risk Factors in Sustainable Water Resources Management. *Water Resources Management*, 37(3). Doi:10.1007/s11269-023-03634-6.
53. Michelsen, N., Dirks, H., Schulz, S., Kempe, S., Al-Saud, M., & Schu, C. (2019). YouTube as a crowd-generated water level archive. *Science Total Environment*, 568: 189-195. <https://doi.org/10.1016/j.scitotenv.2019.05.211>
54. Mohd, Z., Hoque, J., Aziz, N. A., Alelyani, S., Mohana, M. & Hosain, M. (2022). Improving water quality index prediction using regression learning models. *International Journal of Environmental Research & Public Health*, 19 (20): 13702. <https://doi.org/10.3390/ijerph192013702>.
55. Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N. & Abdel-Fattah, M. K. (2022). Prediction of irrigation water quality indices based on machine learning and regression models. *Applied Water Science*, 10-17. <https://doi.org/10.1007/s13201-022-01590-x>.
56. Morton, R. & Henderson, B. L. (2008). Estimation of nonlinear trends in water quality: an improved approach using generalized additive models. *Water Resources Research*, 2-6. <https://doi.org/10.1029/2007wr006191>.
57. Nair, M. R., Ramya, G. R., & Sivakumar, P. B. (2019). Usage and analysis of Twitter during Chennai flood towards disaster management. *Procedia Computer Science*, 115: 350-358. <https://doi.org/10.1016/j.procs.09.089>.
58. Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J. & Liu, J. F. (2020). Streamflow forecasting using extreme gradient boosting model coupled with Gaussian mixture model. *Journal of Hydrology*, 586:124901.
59. Nina, S. & Galina, M. (2021). Digital Techniques For Water Pollution Control. *Journal of Engineering Science*, XXVIII(2), 138-148. [https://doi.org/10.52326/jes.utm.2021.28\(2\).12](https://doi.org/10.52326/jes.utm.2021.28(2).12).
60. Osman, A. A. A., Ahmed, A. N., Chow, M. F., Huang, Y. F. & El-Shafe, A. (2021). Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Engineering Journal*, 12:1545-1556.
61. Panigrahi, B. K., Das, S., Nath, T. K., & Senapati, M. R. (2018). An application of data mining techniques for flood forecasting: application in rivers Daya and Bhargavi, India. *Journal of Industrial Engineering*, Springer; 99:331-42. <https://doi.org/10.1007/s40031-018-0333-9>.
62. Panigrahi, B. K., Nath, T. K., & Senapati, M. R. (2019). An application of local linear radial basis function neural network for flood prediction. *Journal of Management Analysis [Taylor & Francis]*, 6:67-87. <https://doi.org/10.1080/23270012.2019.1566033>.
63. Poursaeid, M. (2023). An Optimized Extreme Learning Machine by Evolutionary Computation for River Flow Prediction and Simulation of Water Pollution in Colorado River Basin, USA. *Expert Systems with Applications*, 233(12):120998. Doi:10.1016/j.eswa.2023.120998.
64. Poursaeid, M. & Poursaeed, A. H. (2023). A new strategy for prediction of water qualitative and quantitative parameters by deep learning-based models with determination of modelling uncertainties. Doi:10.1080/02626667.2023.2293755.
65. Restrepo-Estrada, C., de Andrade, S. C., Abe, N., Fava, M. C., Mendiondo, E. M., & de-Albuquerque, J. P. (2019). Geosocial media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. *Computing & Geosciences*, 111: 148-158. <https://doi.org/10.1016/j.cageo.2019.10.010>.
66. Rosser, J. F., & Leibovici, D. G. (2019). Jackson MJ. Rapid flood inundation mapping using social media, remote sensing and topographic data. *Nat Hazards*, 87: 103-120. <https://doi.org/10.1007/s11069-017-2755-0>.
67. Schapire, R. E. (2003). The boosting approach to machine learning: an overview. In: Denison DD, Hansen MH, Holmes CC, Mallick B, Yu B (eds) *Nonlinear estimation and classification*. Springer, New York, pp 149-171.
68. 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>.
69. Shin, Y., Kim, T., Hong, S., Lee, S., Lee, E., Hong, S., Lee, C., Kim, T., Park, M. S., Park, J. & Heo, T. Y. (2020). Prediction of chlorophyll-a concentrations in the Nakdong River using machine learning methods. *Water* 12(6):1822. <https://doi.org/10.3390/w12061822>.
70. Singha, S., Pasupuleti, S., Singha, S. S., Singh, R. & Kumar, S. (2021). Prediction of groundwater quality using efficient machine learning technique. *Chemosphere*, 276:130265. <https://doi.org/10.1016/j.chemosphere.2021.130265>.
71. Smith, L., Liang, Q., James, P., & Lin, W. (2019). Assessing the utility of social media as a data source for flood risk management using a real-time modelling framework. *Journal of Flood & Risk Management*, 10: 370–380. <https://doi.org/10.1111/jfr3.12154>.
72. Starkey, E., Parkin, G., Birkinshaw, S., Large, A., Quinn, P., & Gibson, C. (2019). Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation. *Journal of Hydrology*, 548: 801-817. <https://doi.org/10.1016/j.jhydrol.2019.03.019>.
73. Sulaeman, D., Arif, S. & Sudarmadji, S. (2018). Trash-polluted irrigation: characteristics and impact on agriculture. *IOP Conference Service Earth, Environmental Science*, 148:012028. <https://doi.org/10.1088/1755-1315/148/1/012028>.
74. Than, N. H., Ly, C. D., Tat, P. V. & Thanh, N. N. (2016). Application of a neural network technique for prediction of the water quality index in the Dong Nai River, Vietnam. *Journal of Environmental Science & Engineering*, 5:7. <https://doi.org/10.17265/2162-5263/2016.07.007>.
75. Tiyasha, T. T. M. & Yaseen, Z. M. (2021). Deep learning for prediction of water quality index classification tropical catchment environmental assessment. *Nat Resour Res* 30(6):4235–4254. <https://doi.org/10.1007/s11053-021-09922-5>.
76. Tuan, N. V. (2020). Regression modelling and scientific discovery. General Publishing House, Ho Chi Minh City.
77. Wu, T., Zhang, W., Jiao, X., Guo, W. & Hamoud, Y. A. (2020). Comparison of five boosting-based models for estimating daily reference evapotranspiration with limited meteorological variables. *PLoS ONE* 15(6):e0235324. <https://doi.org/10.1371/journal.pone.0235324>.
78. Yadav, R. K., Jha, A. & Choudhary, A. (2021). IoT-based prediction of water quality index for farm irrigation. 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, pp. 1443-1448, doi: 10.1109/ICAIS50930.2021.9395921.
79. Yang, G. & Moyer, D. L. (2020). Estimation of nonlinear water-quality trends in high-frequency monitoring data. *Science Total Environment*, 715:136686. <https://doi.org/10.1016/j.scitotenv>.
80. Ye, Q., Yang, X., Chen, C. & Wang, J. (2019). River water quality parameters prediction method based on LSTM-RNN model. In: 2019 Chinese Control and Decision Conference (CCDC), pp 3024-3028.