The Assessment of Water Quality Forecasting Using AI-Based ML Algorithms

Dr. L. Vaikunta Rao[1], Dr. Pravin R.K. shirsagar[2*], Dr. Kuan Tak Tan[3], Dr. Sivaneasan Bala Krishnan[4], Dr. Shrikant V. Sonekar[5], Sayali Zade[6]

[1]*Professor, HOD- Chemistry & Dean R&D,J.B.Institute of Engineering and technology, Moinabad, Hyderabad, India*
[2]*Professor & Dean(R&D), Department of Electronics & Telecommunication Engineering, J D College of Engineering & Management, Nagpur, India*
[3]*Associate Professor and Programme Leader-Engineering Cluster, Singapore Institute of Technology, Singapore*
[4]*Associate Professor & Deputy Director, SIT Teaching and Learning Academy), Nanyang Technological University, Singapore*
[5] *Professor & Principal, Department of Computer Science & Engineering, J D College of Engineering & Management, Nagpur, India*
[6] *Assistant Professor, Department of Electronics & Telecommunication Engineering, J D College of Engineering & Management, Nagpur, India*
Email: pravinrk88@yahoo.com

Abstract: Water quality matters to people, animals, plants, ecosystems, and entities. Recent environmental damage and contamination have harmed water purity. Because they indicate water authenticity, the Water Quality Index (WQI) and Water Quality Classification (WQC) are difficult to predict. In this paper, KNN imputers improve many MLalgorithms for water quality prediction. The precision of current methods is not good enough. Additionally, there are missing values in the dataset that are currently accessible for study, and these missing values significantly impact the classifiers' performance. This paper proposes an automated water-quality prediction system that effectively handles missing data while achieving high forecast accuracy. Furthermore, the accuracy of the suggested approach is assessed in relation to that of four machine learning methods. BPNN, RBFNN, SVM, and K-Nearest Neighbours are these approaches. These approaches model and predict water quality parameters such as DO, pH, NH3, NO3, and NO2. For the purpose of evaluating the accuracy of the various approaches to prediction, published data were utilized and for DO prediction, BPNN, RBFNN, SVM, and KNN had Pearson correlation values of 0.60, 0.99, 0.99, and 0.99. BPNN, RBFNN, SVM, and LSSVM had Pearson correlation coefficients of 0.56, 0.84, 0.99, and 0.57 for pH prediction. For NH3-N forecasting, BPNN, RBFNN, SVM, and LSSVM had Pearson correlation coefficients of 0.28, 0.88, 0.99, and 0.25. For NO3-N prediction, BPNN, RBFNN, SVM, and LSSVM obtained coefficients of correlation of 0.96, 0.87, 0.99, and 0.87. With correlation ratings of 0.87, 0.08, 0.99, and 0.75, BPNN, RBFNN, SVM, and LSSVM projected NO2-Ncorrelated coefficients. SVM was used to forecast water quality for groundwater-based industrial aquaculture systems. Most precise and reliable predictions were made with SVM. Both published and commercial aqua farming system data showed that the support vector machine (SVM) had the highest prediction accuracy, with 99% accuracy. For commercial farming water quality modeling and forecasting, utilize the SVM model.est.

Keywords: water-quality prediction, WQI, KNN imputer, ML Algorithms, RBFNN, BPNN, SVM, KNN.

1. Introduction

All life depends on water, which is among the most valuable resources. Contamination lowers the water's overall quality, impacts the well-being of marine life, and ultimately impacts the people who utilize it. Therefore, monitoring water quality and ensuring the survival of aquatic animals are crucial. Understanding the problems and concerns related to water quality is also necessary for managing and minimizing water pollution. In an attempt to comprehend the condition of the maritime environment, some governments worldwide have begun to create environmentally controlled water programs. Water management and quality control must improve to ensure safe, affordable drinking water [1]. Systematic clean water, garbage management, and operational monitoring assessments are needed to overcome these difficulties. Water quality prediction occurs when water quality changes are predicted. Planning and regulating water quality requires assessment results. Plans for water diversion should evaluate the general homogeneity of the water. Research on methods to predict water quality today is necessary since a significant amount of water is used to address routine drinking issues. Security concerns require AI and ML to recognize the relationship between system inputs and outputs[2,9]. This is because it eliminates the need to rely on complicated procedures. Forecasting prospective modifications to water quality at various degrees of pollution and developing realistic solutions for avoiding and regulating water contamination are two ways in which prevention and regulation approaches can be enhanced. Water pollution studies must incorporate water quality forecasts to evaluate water ecological protection. It's essential for water planning, regulation, and accountability. A practical and realistic water quality forecast is crucial. Future water quality must be predicted to prevent rapid changes and suggest solutions. Standard water quality forecasting ignores biology, physics, hydraulics, the science of chemistry, and meteorological data [3].

Aquatic ecosystems and human health have suffered due to pollutants and pollution that have recently degraded the water's quality. This metric simplifies water quality comparisons and interpretation across numerous places and periods of history by reducing complexity to a single value. Among the many biological, chemical, and physical elements that WQI considers are turbidity, dissolved oxygen, pH, nutrition levels, and pollutants.WQI evaluates water quality thoroughly to aid water management decisions by incorporating these aspects. Additional functions include water quality classification (WQC), which classifies water samples by thresholds [2]. This classification offers a helpful framework for calculating water contamination levels, allowing for targeted interventions and legislative actions. By grading water quality, stakeholders can identify issues, prioritize correction measures, and protect water resources. The need to address the decrease in water quality and its impacts motivated the probe. Long-term development, public health, and ecosystems are all seriously threatened by water pollution and contamination.To identify potential issues, implement efficient management strategies, and ensure that various industries have access to safe, healthy water, water quality tracking and evaluation are crucial. Machine learning and AI have changed water quality management.The advancements above caused a substantial change in the methods used to analyze and interpret water quality data. Machine learning algorithms are highly beneficial in various domains, including predicting contamination episodes, determining the source of contaminants, and enhancing treatment methods. The capacity of these algorithms to effectively manage enormous amounts of data and spot complex patterns allows academics to make substantial progress in several different areas of study. AI-powered solutions are increasingly being used in the field of managing the quality of water, which includes not just monitoring but also anticipating and responding to such situations [4,11]. By offering proactive water resource monitoring, water resource management systems facilitate the early detection of possible problems, mitigating environmental and public hazard.

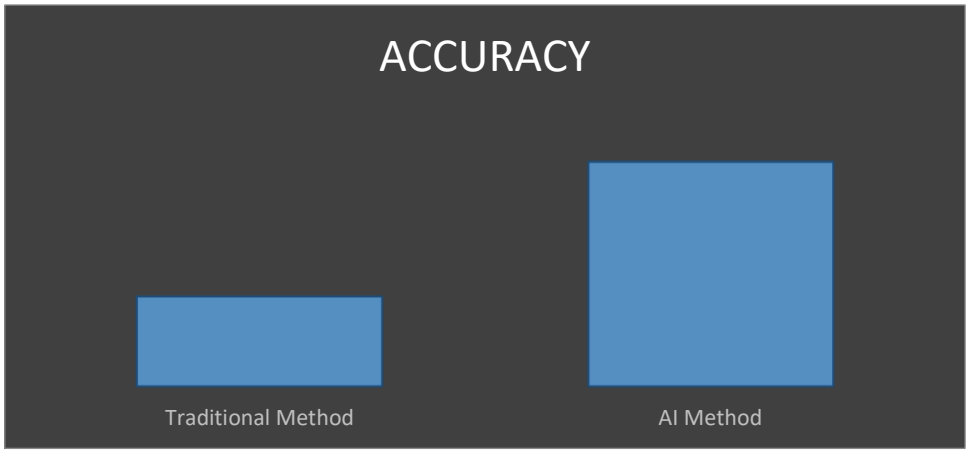

Fig 1: Traditional vs. AI-enabled water quality system monitoring adaptation

Figure 1 depicts how AI-based water quality monitoring and treatment is approaching a turning point. Innovative uses of artificial intelligence (AI) include neural networks for infection and metal detection, algal bloom prediction models, and water treatment automation using AI-powered decision-making tools. Experts are studying data privacy, security, and ethical AI for environmental management. AI water quality restoration and monitoring uses technology. It also embodies a commitment to environmental sustainability and public health, indicating a future where ecology and technology coexist [5, 7]. In this paper, machine learning predicts WQC and WQI. The following are the contributions of this paper:

- Data preparation methods, such as data reduction (mean imputation) and data standardization, were employed to fit the data and facilitate further processing.
- WQI is predicted using four regression models, and WQC parameters are optimized and adjusted using grid search utilizing four categorization models.
- Four metrics were used to analyze the regression models' performance: square MSE, MAE, and R2. Precision, accuracy, recall, and F1 score measured classification efficiency.
- When it came to the classification of WQC, compared to other models, the SVM model fared best.

2. OBJECTIVES

By identifying areas or circumstances where water quality may decrease before serious problems arise, machine learning techniques such as RBFNN, BPNN, SVM, and KNN can be used to enhance prediction models that can accurately evaluate water quality and recognize possible contaminants before they lead to major issues. This is the main goal of applying machine learning (ML) to forecast water quality. By doing so, proactive measures can be taken for avoiding pollution, control water resources successfully, and safeguard public health.

3. BACKGROUND

Water quality is very important for both human health and the environment. Key water quality indicators include Dissolved Oxygen (DO), pH, and nitrogen compounds like NH3-N, NO3-Nand NO2-N. These indicators tellabout the health of water bodies, such as lakes, rivers, and oceans.

- Dissolved Oxygen (DO): This tells us how much oxygen is in the water. Fish and other aquatic life need oxygen to survive. Low DO levels can harm aquatic life.
- pH: The pH level shows if the water is acidic or basic. If the pH is too high or low, it can harm aquatic organisms and affect the overall health of the ecosystem.
- Ammonia Nitrogen (NH3-N): Ammonia is toxic to aquatic life, and it usually comes from waste, fertilizers, and pollution.
- Nitrate Nitrogen (NO3-N) and Nitrite Nitrogen (NO_2-N): High levels of these nitrogen compounds can lead to harmful algae blooms and reduce oxygen in water, which can harm aquatic ecosystems.

Traditionally, measuring these water quality parameters involves collecting water samples and sending them to a lab for analysis, which takes time and is costly. With advances in technology, we now have

sensors and other tools to collect a lot of data on water quality. This opens up the possibility to use ML algorithm to predict water quality in real-time, helping us manage water resources more effectively.
Machine learning (ML) can help make predictions about water quality by learning from past data. In this paper, we focus on using four popular machine learning algorithms listed below. These algorithms are effective at finding patterns in complex data and can predict water quality parameters based on historical data.

- RBFNN: This type of neural network is good at finding non-linear relationships in data, making it useful for predicting water quality based on multiple factors.
- BPNN: A common neural network used for learning from data and improving predictions by adjusting itself over time. It is effective for predicting things like DO, pH, and nitrogen levels.
- SVM: A machine learning algorithm that finds the best boundary (or line) that divides data into different categories. It's used to predict water quality based on data patterns.
- KNN: A simple algorithm that makes predictions based on the "neighbors" (or nearby data points) of a given point. For water quality, it predicts the values based on past similar situations.

By using these algorithms, we can predict water quality parameters like DO, pH, NH_3-N, NO_3-N, and NO_2-N without the need for time-consuming and expensive lab tests. However, there are still challenges. For example, data may not always be perfect, and models may not always work well in different areas or conditions. Also, it can be difficult to understand how some of these machine learning models come up with their predictions.
This paper aims to explore how well these machine learning algorithms work for predicting water quality. We will review their strengths and limitations and discuss how they can help improve water quality management and monitoring.

4. LITERATURE REVIEW

B. Amirgaliyevet al. [1]introduced an IoT-based predicted quality control system. Enable data transfer via Bluetooth or Wi-Fi networks. The project's scientific originality is creating a revolutionary system that integrates cutting-edge IoT technology and machine learning techniques to offer thorough and precise water quality monitoringa significant contribution to environmental safety and water management.
Karumanchiet al. [2] proposed a technique using sensor devices to gather data on water quality factors such as turbidity, pH, temperature, and dissolved oxygen. A dataset is created Using these criteria, and machine learning (ML) techniques are used to assess processing speed, accuracy, precision, and recall. To determine the primary factors affecting E. coli levels, our machine learning algorithms also create a matrix of correlations among water quality data. We used a variety of machine learning approaches on the dataset, such as ensemble techniques that combine each of these algorithms, Random Forest Classification (RFC), Support Vector Regression (SVR), and XGBoost.
Shafi et al. [3]implemented a different computerized system that can forecast the type of water that exists based on various target classes, which was the paper's main goal. These classes include irrigation water, drinking water, and outdoor bathing water. Conductivity and pH were used to calculate these classes. The Indian Government's website has officially provided raw data on several Indian regions over the last five to six years, containing fecal coliform, dissolved oxygen, pH level, nitrate, and conductivity. To further enhance the classification performance, it was preprocessed, scaled, and enhanced using the KNN, feature scaling, and SMOTE techniques, respectively.
Shams et al. [5]discussed the accuracy of both the WQC and WQI, a key indication of water validity. We use machine learning to forecast WQI and WQC, including improving and optimizing parameters to improve several machine learning models. One essential technique for adjusting and fine-tuning parameters for four regression models and four classification models is grid search.
Joshi et al. [6]suggested natural language processing techniques like stemming, lemmatization, and stop-word removal, including Part-of-Speech (POS) tagging are used to process textual opinions. Predicting user preferences and comprehending their cognitive processes are the goals of a recommendation system. The system provides customized information by taking user preferences and needs into account. A more thorough analysis of the data is required to improve the quality of recommendations.
Grekov et.al [13]addressed Chernaya River data from Sevastopol, Crimea. Four standard unsupervised machine learning methods, iForest, elliptic envelope, one-class SVM, and local outlier factor, were applied to recognize bivalves' warning signs.With an F1 score of 1, the findings demonstrated that

abnormalities in mollusk activity information could be identified using the elliptic envelopes, iForest, and LOF approaches with appropriate hyperparameter tweaking without generating false alarms. The iForest approach is the most effective, according to a study of anomaly detection times. These results show that bivalve mollusks can be used as biological indicators in automated systems of monitoring to detect pollution in aquatic ecosystems early.
Karras et al. [14]discussed the functioning and implementation of a Geographic Information System (GIS) and a specialized expertise system examined to monitor and record the development of dangerous illnesses on fish farms. Targeting Greece's aquaculture regions specifically, the technology gathers climate and topographical information relevant to these farms. This system's capacity to determine the times between separate cages and larger fish farm entities is one of its features; It provides essential details about the mechanics of disease transmission. This information then serves as a starting point for our expert system

5. AI FOR WATER QUALITY MANAGEMENT

Chemical, mathematical, and statistical data from water quality models is significant. Because so much data is collected from different water bodies, models and predictions are essential for water quality measures. Artificial intelligence is necessary to study and forecast water quality models [6]. Figure 2 shows how artificial intelligence models analyze the architecture for water quality monitoring and device control

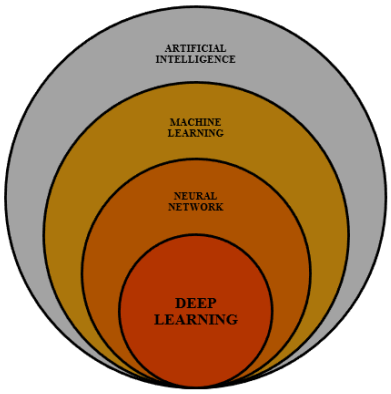

Fig 2: A simplified visualization shows the link between deep learning and artificial intelligence.
Rivers, lakes, reservoirs, and oceans can all have their surface water quality monitored and evaluated using AI models. Among other essential benefits for water quality assessment and monitoring, artificial intelligence (AI) is quick, effective, affordable, and capable of real-time surveillance of water quality and prediction. Recently published research suggests machine learning can be used to analyze surface water quality.By gathering data, training models, choosing the right algorithms, and using AI models, surface water quality in lakes, rivers, seas, and water reservoirs may be tracked and assessed [7,10]. AI has numerous significant advantages for water quality monitoring and evaluation, such as being rapid, efficient, cost-effective, and able to be employed for real-time water quality monitoring and prediction. Machine learning techniques help assess the quality of ground waters, according to current studies and literature reviews. Data gathering, model training, model validation, and method selection must all be finished before machine learning is used.

6. WATER QUALITY INDEX

Using 10 of the most commonly used water quality attributes, alkalinity, chloride, coliforms, specifically conductivity, pH, and oxygen dissolution (DO), Horton (1965) created WQI in the US. It has been widely adopted and used by nations across Asia, Africa, and Europe. The given weight reflects the importance of a property for a specific usage and significantly affects the index. Recently, many scientists and specialists have thought about changing the WQI notion in a few different ways [8, 12].
The WQI grade considers the cumulative effects of different water quality indicators. This strategy is one of the finest ways to inform decision-makers and concerned citizens about water quality. Finding potential uses for water bodies, managing them responsibly, and classifying them according to their biological, chemical, and physical characteristics are the goals of the WQI approaches. The components

are assembled as illustrated in Figure 3, reliable criteria are applied, and each parameter is given the proper weight in WQI approaches, which can be viewed as models for evaluating WQ. All WQI techniques use four common steps to complete their computations.

•Choosing the necessary variables;

•these variables are then changed using a comparable scale after having different dimensions.

•By assigning a weighting factor to every converted variable, subindices are created.

•The process by which subindices are combined to get the result index score. The part of the guide that follows provides information on the evolution of WQ indicators and an overview of the primary indicators used worldwide to evaluate WQ.

Water quality metrics are classified into four fundamental categories.

•Public indices: NSF WQI evaluations employ comprehensive water quality indicators, not usage.

•Consumption indices: These indexes categorize water according to its use and application (industrial, drinking, ecosystem preservation, etc).

•Design and planning indices: These are instruments to assist in WQ management decision-making and project planning projects.

•Statistical indices: These are based on statistical approaches rather than individual views. Statistical methods are implemented to evaluate the data in the present instance. The statistical substantiation of specific predictions related to WQ based on observation is another critical stage of the statistical method [9].

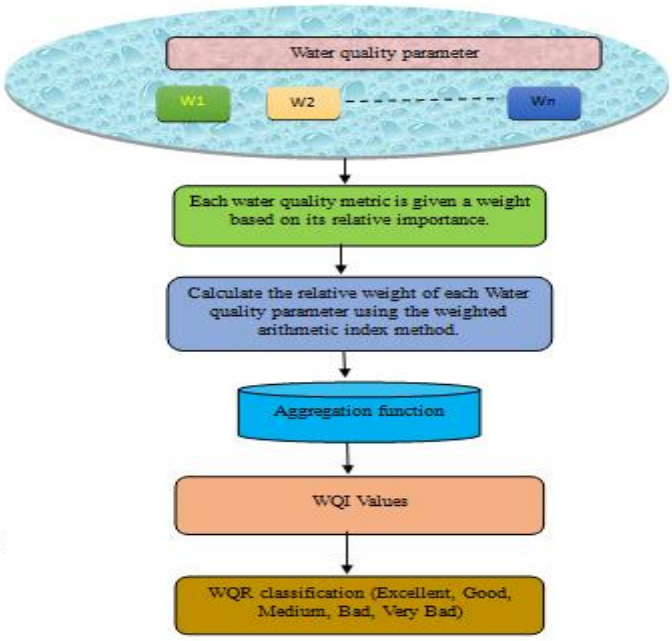

Fig. 3: Basic water quality index development techniques

7. A SELECTION OF THE MOST COMMONLY USED WQI TECHNIQUES

Numerous domestic and international organizations have worked to develop a collection of water quality measures over the years, which have been used to evaluate the WQ in various contexts. A single index cannot sufficiently capture the entire WQ in that waters on a global scale. WQI must be constructed to measure changes in WQ across time and space and to evaluate the objectives of global agreements aimed at conserving water resources [9, 12]. The following are some standard indices used by various WQI approaches:

NSF WATER QUALITY INDEX

"NSF-WQI" refers to the designation associated with the WQI that was approved by the National Sanitation Organization. An NSF-WQI has the most significant degree of clearance for use in evaluating the purity of the country's water despite the fact that it has been widely criticized for not adequately evaluating all US zones [23, 28]. The Dalkey technique was the foundation for this method's precise parameter selection. After that, they developed a standard scale with appropriate weights for each [10]. The WQ is measured using nine metrics: turbidity, temperature, pH, DO, BOD, PO4, NO3, fecal

coliform bacteria & TDC. A weighted curve graph converts the WQI to a chance score Q_i. Table 1 displays the relative values of each parameter based on the NSF-WQI [9].
$\mathrm{NSF-WQI} = \sum_i^k Q_i W_i$ (1)
where k represents how many parameters there are.
Q_i −classification of the jth parameter's quality;
W_i −weight with relation to the jth parameter
$(\sum W_i = 1)$.Table 2 displays the water quality rating as determined by the NSF-WQI.

Table 1: Weights and Parameters for the NSF WQI technique parameters

Parameter	Weight	Parameter	Weight
DO	0.17	PO_4	0.10
Fecal coliforms	0.16	Temperature	0.10
pH	0.11	Turbidity	0.08
BOD	0.11	TDS	0.07
NO_3	0.10	Total	1.00

Table 2: Water Quality Rating Based on NSF-WQI

NSF-WQI	WQR
90-100	Excellent
70-90	Good
50-70	Medium
25-50	Bad
0-25	Very bad

This method has the advantage of objectively, quickly, and reproducibly summarizing the information regarding the importance of a single indicator. Furthermore, it is easy enough for average people to use and not only for experts [11].

TheCANADIAN COUNCIL OF ENVIRONMENTAL MINISTERS WATER QUALITY INDEX

The Canadian Council of Environmental Ministersintroduced a comprehensive assessment of river systems' functionality for maintaining aquatic life at specific management points across Canada. The CCME-WQI connected water quality data to several advantageous water applications by using pertinent WQ criteria for reference points. The index is created over a specified period for each monitoring location [26, 27].

Numerous water quality characteristics are analyzed using water samples gathered within a given time frame. The applicable water quality standard is compared to all measured parameters [13]. The three criteria used to calculate the index are the percentage of variables and testing thatthat do not adhere to the rules, along with the difference between the permissible norms for tests that fail to satisfy the requirements. Equation 2 defines these three elements, which are scope (S), the frequency (F), the amplitude (A).

$$S = \frac{\mathrm{Numberofincorrectaspects}}{\mathrm{total\ Number\ of\ aspects}} \times 100 \qquad (2)$$

Scope (S) is the proportion of failing parameters to all parameters that, at a specific time frame, are inconsistent with water's quality standards at any particular location, as indicated by Equation 3.

$$F = \frac{\mathrm{Numberoftestsfailed}}{\mathrm{total\ Number\ of\ tests}} \times 100 \qquad (3)$$

where frequency (F) represents the proportion of each test that does not meet water quality standards. The test is deemed unsuccessful if any sample parameter value exceeds the specified limit. The total quantity of failed trials over the chosen period includes the parameters of each failed sample [12]. To find the entire number of tests for a particular location, multiply the number of specimens measured at the chosen times by the number of mean parameters assigned to each sample [14].

Amplitude (A): Equation 4 shows how the amplitude component is the average deviation of refused outcomes from their norms.

$$A = \frac{\mathrm{nse}}{0.01\ \mathrm{nse} + 0.01} (4)$$

Compute the excursion, which is the test result's relative difference from its typical value:
When the test's results fall within the range that is considered acceptable by Equation 5:

$$\text{excursion}_v = \left(\frac{\text{failed test value}_v}{\text{standard value}_v}\right) - 1 \quad (5)$$

When the test's result shouldn't fall shorter of the standards stated in Equation 6:

$$\text{excursion}_v = \left(\frac{\text{standard value}_v}{\text{failed test value}_v}\right) - 1 \quad (6)$$

Using the following expression 7, the total quantity of tests that are not in concordance is calculated:

$$\text{nse} = \frac{\sum \text{excursion}_v}{\text{total number of tests}} (7)$$

In this case, one represents the total of all the normalized deviations over the norms. The following method is used to calculate the CCME-WQI using Equation 8 [14]:

$$\text{CCME} - \text{WQI} = 100 - \left(\frac{\sqrt{S^2 + F^2 + A^2}}{1.732}\right) (8)$$

The following factor, 1. 732, is used to transform the data acquired from the CCME-WQI technique to an index of 0-100. The previous method provides a digital value for the water quality in addition to producing a CCME-WQI result. A score of 0 denotes extremely poor water quality, while a score of 100 denotes excellent water quality. Table 3 suggests the CCME-WQI based upon WQR [15].

Table 3: WQR as per CCME-WQI

CCME	WQI
95-100	Excellent
80-94	Good
60-74	Fair
45-59	Marginal
0-44	Poor

The OREGON THE WATER QUALITY INDEX

The O-WQI assesses and expresses eight distinct WQ components. This technique examines several parameters, including temperature, nitrate, and ammonia, overall solids, nitrogen and phosphorus, pH level, dissolved oxygen, biological oxygen demand, and fecal coliform.

In order to get an agreement on the most recent understanding regarding how to handle a difficult situation, this strategy might be characterized as obtaining information from multiple specialists [16]. Using logarithmic procedures, the outcomes of several water quality parameters were converted into subindex numbers for both indices [17]. Logarithmic transforms have the advantage of having a more significant effect on modifications to magnitude at lower impairment levels than at higher impairment levels. The subsequent Equation 9 is supplied by

$$O - \text{WQI} = \sqrt{\frac{n}{\sum_{k=1}^{n} \frac{1}{STI_k^2}}} (9)$$

Each parameter's subindex is denoted by STI, and n is the number of subindices [17].

Table 4: WQR based on O-WQI

Q-WQI Value	WQR
90-100	Excellent
85-89	Good
80-84	Fair
60-79	Poor
0-59	Very poor

The consolidation technique employed for integrating the sub-indices enumerated in Table 4 has the advantage of granting the most impacted aspects on the final WQI. Furthermore, the process is highly susceptible to fluctuations in environmental conditions and significantly influences water quality [21, 24].

8. PROPOSED APPROACH FOR WATER QUALITY ASSESSMENT THE MOST COMMONLY

One of the biggest environmental problems civilizations are currently dealing with is water contamination, and the harm it causes is primarily caused by a lack of emergency management, alert systems, and forecasting abilities. Therefore, in order to promote responsible decision-making and water quality regulation, an appropriate monitoring and early notification system must be put in place immediately [9,18]. The suggested water quality prediction method is depicted in Figure 4. In recent years, many machine learning algorithms have advanced.A machine learning system assesses water quality using oxygen solubility, pH, conductivity, biological oxygen requirements, nitrates, fecal coliform, and total coliform [21]. Datanormalization and mean imputation are examples of preparation already done on the dataset.

Description of Dataset

Information was gathered from an article about aquaculture pond monitoring and data from a typical industrialized aquaculture operation using groundwater. The temperature, DO, pH, NH 3-N, NO3-N, and NO2-N were among the data collected [19, 23]. Groundwater has a pH of 7.8. The dilute water was reused for this system after the groundwater supply was diluted 1:1 with tap water. Daily water samples were taken from the industrialized aquaculture system, and a monitoring system created in the authors' lab was used to measure the water quality parameters. Table 5 provides a list of the specific measurement techniques. DO, pH, NH 3-N, NO3-N, & NO2-N were among the measured parameters.

Table 5: Techniques for measuring every aspect of water quality.

Data	Measurement Methods
DO	DO sensor
pH	pH meter
NH3-N	Nessler's reagent spectrophotometry
NO3-N	Ultraviolet spectrophotometric method

The tilapia was fed in a recirculating fashion utilizing the industrial aquaculture method. With a feed supply of 5% of the fish weight, the tilapia density was maintained at 600 g/m3. Throughout the cultivation phase, the intake water was recycled.The following stages were used in the model screening technique for this investigation

. Initially, water quality was predicted and simulated using the four models and the data. After that, the best model had been chosen based on the forecasts' results [10,19]. The most effective approach was then utilized to predict and simulate water quality using data collected from a genuine industrial aqua-farming business that uses water from the ground as its primary supply to ensure the model's relevance [28].n Figure 5. In order to find any noteworthy correlations or dependencies between the variables, the correlation matrix investigates the connectionsbetween the various properties.

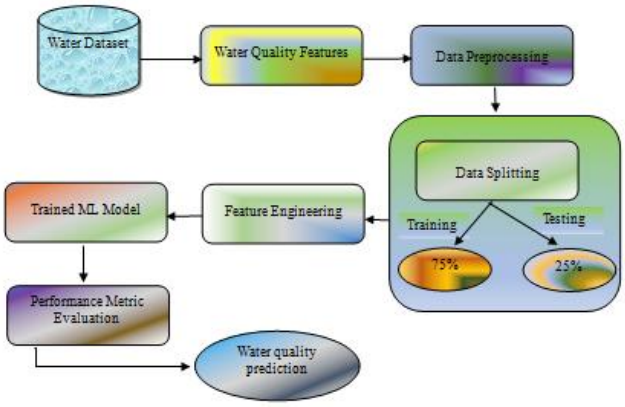

Fig 5:Heat map visualization of the feature correlations

between the various used used in the model screening technique for this investigation. Initially, water quality was predicted and simulated using the four models and the data. After that, the best model had been chosen based on the forecasts' results [10,19]. The most effective approach was then utilized to

predict and simulate water quality using data collected from a genuine industrial aqua-farming business that uses water from the ground as its primary supply to ensure the model's relevance [28]. n Figure 5. In order to find any noteworthy correlations or dependencies between the variables, the correlation matrix investigates the connections between the various properties.

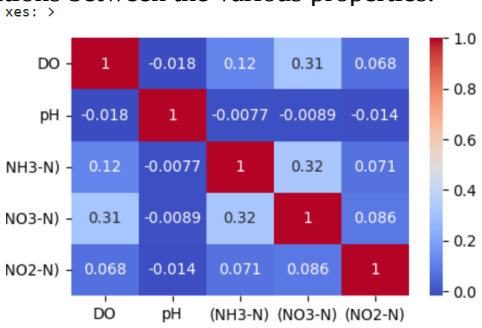

Fig 5: Heat map visualization of the feature correlations

8.2 Preprocessing of data

Data analysis requires processing to improve quality. This phase calculated WQI using the dataset's most critical parameters. WQI classed water samples. Normalizing with z-scores improved accuracy [20].

8.2.1 KNN Imputer

KNN imputer finds nearest neighbors using Euclidean distance matrices, making missing values easy to impute. Euclidean distance is calculated by deleting missing values and improving non-missing coordinates. Equation 10 and 11 are used to calculate Euclidean distance [26].

$$Q_{ab} = \sqrt{\text{weight} \times \text{squared distance from present coordinates}} (10)$$

Where

$$\text{weight} = \frac{\text{Total count of coordinates}}{\text{count of present coordinates}} (11)$$

8.2.2 Dataset Missing Values Elimination

Data processing technique two involves removing missing values. The next series of trials employs this technique, eliminating any fields that have missing data [21].

8.2.3 How to calculate the WQI

Its water quality index is one of the primary indicators of water quality. WQI is calculated based on several characteristics. WQI may be determined using Equation 12:

$$\text{WQI} = \frac{\sum_{t=1}^{P} q_t \times w_t}{\sum_{t=1}^{P} w_t} (12)$$

where P is the total quantity of the parameters, w_t is the variable t's weight in units, and q_t is the variable t's quality scale. $q_t$is calculated using Equation 13:

$$q_t = 100 \times \left(\frac{v_t - v_{id}}{s_t - v_{id}}\right) (13)$$

where v_{id} is the ideal value for each parameter in the variable's t in the context of pure water, $s_t$is the typical value, and $v_t$is the projected quantity of the parameter t. Equation 14 is used to get the unit of weight, w_t. As shown in Table 6, wherein v_t is the value that is measured about the quantity of water sample tested and s_t is the average value recommended for parameter t, $v_{id}$is an ideal value that denotes water that is pure (0 for each parameter except DO = 14.6 milligrams per, pH = 7.0).

$$w_t = \frac{H}{s_t} (14)$$

Equation 15 is utilized to determine the proportionality constant, denoted by H.

$$H = \frac{1}{\sum_{t=1}^{P} S_t} (15)$$

Table 6: Water Quality Classification (WQC)

Water Quality Index Range	Classification
0-25	Excellent
26-50	Good
51-75	Poor
76-100	Very Poor

Our selection parameters are among the extra factors that can be calculated with WQI. The variable data is used to calculate the WQI [22]. Any parameter may be tested using the suggested system utilizing any data on water quality.

8.2.4 Method of Z-Score Normalisation

Data are normalized using the Z-score, which calculates mean (μ) and standard deviation. The scale parameter values ranging from 0 to 2 were subjected to the Z-score. Equation 16 is used to compute it:

$$\mathrm{Z-score} = \frac{(y - \mu)}{\sigma} (16)$$

8.3 WATER QUALITY PREDICTION ML ALGORITHMS

The machine model of learning possesses strong self-learning and nonlinear fit features. One of the most popular models with distinctive features for machine learning is BPNN. There are no issues with local minimums in BPNN. The drawbacks of BPNN, including overfitting and poor repeatability, can be addressed by SVM [23, 28]. SVM doesn't need a priori architectural specifications and can be generalized for small amounts of data. According to the structural danger minimization approach, SVM is better suited for predicting water quality measurements since it can maintain prediction accuracy and operational efficiency while processing data with unforeseen changes. In this work, RBFNN & KNN were also utilized for approach pre-screening [24]. The pretest results showed that SVM's accuracy was much higher than that of RBFNN, BPNN, and KNN, which were below 40%. As a result, RBFNN, BPNN, and KNN are not employed in subsequent research. The RBFNN, BPNN, SVM, and KNN models were ultimately selected for this study's water quality prediction.

8.3.1 Back Propagation Neuron Network (BPNN)

One popular neural network technique used in an intelligence data processing system is BPNN [23]. The error-correcting back-propagation method is used to train BPNN input data. To determine the optimal prediction value, BPNN, a multilayer feed-back network, employs gradient descent. Standard BPNNs have three layers: input, hidden, and output. BPNN employs a sigmoid nonlinear function, and Equation 17 serves as its transfer function [25].

$$f(y) = \frac{1}{1 + e^{-y}} (17)$$

8.3.2 RBFNN, or Radial Basis Function Neurone Network

RBFNN processes human feelings using neural network topology to replicate a three-layer comments system with one hidden layer. It offers a basic network structure, rapid convergence, and easy training [13]. For real-time control purposes, RBFNN is a suitable local approximations network because it prevents local minor issues and speeds up learning. A set of perceptual units makes up the input layer of an RBFNN, an implicit portion of calculation nodes, and a compute node in the output layer. Equation 18 illustrates that the nonlinear conversion function represents the radial base function in the RBF system's hidden layer.

$$L(W_s - D_t) = \exp\left(-\frac{\|W_s - D_t\|^2}{2\sigma^2}\right), t = 1,2, \ldots\ldots, q \qquad (18)$$

W_s stands for the s-th input samples, The t-th centre point is D_t, and the layer has q concealed nodes [26]. The most crucial parameters associated with the radially foundation function are the function center D_t, width σ_t, and layer hiding weights ω_t. Equation 19 explains how the RBF network acquires the result at the output layer via a linear transmission [16]:

$$U_t = \sum_{m=1}^{K} \omega_{mt} \exp\left(-\frac{\|W_s - D_t\|^2}{2\sigma^2}\right), m = 1,2, \ldots\ldots, l \qquad (19)$$

Where l represents the hidden layer weights and the total number of samples that must be produced.

Finding the center of the function's radial basis is the first step. The self-organized center selection approach, a subset of supervised learning known as tutorless learning, is a well-liked algorithm for locating the center [27]. Equation 20 explains this process. Using k-means clustering, this study found each suggested node's center ci:

$$D_t = \frac{1}{l_t}\sum_{W_n \epsilon D_t} W_n \quad (20)$$

W_n represents the k-th cluster center and $l_t$represents the overall count of samples taking part in the training or test. When the group's center shift is smaller than the predetermined constant, clustering comes to an end [17, 18].

The Gaussian function is chosen by the essential function, and Equation 21 can be used to get the width i:

$$\sigma_t = \frac{D_{max}}{\sqrt{2q}}, t = 1, 2, \ldots, q \quad (21)$$

where $D_{max}$is the farthest distance between the selected centers.

After determining the hidden layer node's center and width, gradient descent, least-squares, and pseudo-inverse can construct output weight vectors.

8.3.3 SVM algorithm

The SVM's primary goal is identifying the optimum classification surface for each training sample. Assume the sample set is (W_t,U_t), t=1, 2..., l. $W_t\left(W_t \epsilon L^l\right)$, is the input number of the i sample, and $U_t \epsilon L^l$ is the equivalent output value. Equation 22 defines high-dimensional feature spaces and linear regression functions.

$$f(w) = \omega^T \varphi(w) + c \quad (22)$$

with f(x) the permissible deviation interval, y the predicted value, and f(w) the regression function's projected value [28, 29]. A value reduction of zero occurs when the variance between f(w) and u is more minor than.

The main goal of training an SVM model is finding the ideal and b to make f(w) as near to u as feasible. Thus, by adding the relaxed variables $\varepsilon_i^1, \varepsilon_i^2$, it becomes a convex quadratic formula problem:

Finding the ideal and b to get f(w) as near to u as is feasible is the main goal of creating an SVM model. Consequently, a convex quadratic formula problem arises when relaxed variables ε_i^1 ε_i^2 are added:

$$\begin{cases} \min \frac{1}{2}\|\omega\|^2 + C\sum_{t=1}^{l}(\xi_t^1 + \xi_t^2) \\ s.t \begin{cases} U_t - \omega^T(W_t) - c \le \varepsilon + \xi_t^1 \\ -U_t - \omega^T(W_t) + c \le \varepsilon + \xi_t^2 \\ \xi_t^1 \ge 0, \xi_t^2 \ge 0, i = 1,2, \ldots, l \end{cases} \end{cases} \quad (23)$$

However, C serves as a punishment factor; the sample having a training defect more significant than $\boldsymbol{\varepsilon}$ will be penalized more heavily if C is more significant [29 ,30]. The regression function's error decreases as $\boldsymbol{\varepsilon}$ decreases. As stated in Equation 23 above, $\boldsymbol{\varepsilon}$ establishes the error criteria for the regression coefficient.

As seen in Equation 24, the finalized SVR regression function can be obtained by substituting the appropriate kernel function $k(w_t, w)$ for the inner product vectors in highly dimensional space$\varphi(w_t . \varphi(w)$ by introducing the Lagrange function.

$$f(w) = \sum_{t=1}^{1}(\alpha_t - \alpha_t^*)K(w_t, w) + c \quad (24)$$

Where α_t and α_t^* are Lagrange multipliers.

8.3.4 K-Nearest Neighbors (KNN) Model

One of the oldest machine learning methods for data classification is the KNN Model. The KNN algorithms determine the nearest location between the elements by using K-neighbour values. The K-value, a unique number, is utilized to identify the nearest points in the characteristic vectors. Three K-values were chosen for this investigation to provide positive results. Equation 25 illustrates how to use the Euclidean distance algorithm Q_t to find the feature vector's closest neighbor [5,12,30].

$$Q_t = \sqrt{(W_1 - W_2) + (Z_1 - Z_2)^2} \quad (25)$$

Where$W_1, W_2, Z_1,$ and Z_2 are variables for input data.

9. RESULTS AND DISCUSSION

Following the application of supervised learning techniques to the dataset, the classification models' outcomes are calculated using performance metrics. To establish the best methods for distinguishing potable and noteworthy water, this part presents the outcome data and matrices of confused the machine learning algorithms employed in this paper. A confusion matrix with a "1" indicating that the water is consumable and a "0" indicating that it is not is displayed in Table 7.

Table 7: General layout of the confusion matrix

	Predicted (1)	**Predicted (0)**
Actually (1)	t_p	f_p
Actually (0)	f_n	t_n

The following is a reading of the four notations that are shown in a matrix:

- The term t_p is the quantity of recordings that are believed to be drinkable but aren't.
- $f_p$is the percentage of recordings that are really drinkable despite being anticipated to be unpotable.
- $f_n$is the number of records that are expected to be consumable but are revealed to be unfit.
- It is represented as t_n, the number of records that are not consumable even if they are expected to be.

This method shows water potability-based machine learning. This study measured performance using ROC AUC and F1-score. The two elements that determine the F1 score are precision and recall. The percentage of portable samples that the predictive algorithm recovered out of all the samples is known as the precision of an ML classifier. It can be calculated using Equation 26 as follows:

$$\text{Precision} = \frac{t_p}{t_p + f_p} \tag{26}$$

On the other hand, the number of samples that the algorithm for machine learning properly recognized as portable among all the portable samples is known as the recall rate. Equation 27 is utilized in its computation.

$$\text{Recall} = \frac{t_p}{t_p + f_n} \tag{27}$$

Calculating the F1 score can use recall and precision levels. In this example, Equation 28 represents harmonic mean recall and precision.

$$\text{F1} - \text{score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{28}$$

The ROCAUC, an assessment that considers the ability to differentiate between classes, is computed using sensitivity and recall. Equation 29 is the one used to determine sensitivity, whereas (2) already determines recall:

$$\text{Sensitivity} = \frac{f_p}{f_p + t_n} \tag{29}$$

The ROC is the ratio of the number of accurate forecasts (Recall) for the positive class to the proportion of errors (Sensitivity) to obtain the negative class. The proportion of accurately anticipated cases is known as accuracy. It can be computed with Equation 30:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{30}$$

Regression models were evaluated using mean square error, MAE, and R2. To compute MAE, use Equation 31.

$$\text{MAE} = \frac{1}{\text{P}} \sum_{t=1}^{\text{P}} \left| \text{Z}_{\text{real}_t} - \text{Z}_{\text{pred}_t} \right| \tag{31}$$

The MSE is computed using Equation 32.

$$MSE = \frac{1}{P}\sum_{t=1}^{P}(Z_{real_t} - Z_{pred_t})^2 \quad (32)$$

Machine-learning water quality forecasting model due to its reliability and versatility. Due to its many applications and good results, RBFNN was employed to forecast water quality. BPNN has a more substantial generalization capability. BPNN and RBFNN were utilized to forecast water quality in this investigation because RBFNN generalizes better. SVM's strong generalization and approximation skills enable it to get around problems that neural networks find difficult to avoid. One drawback is that the SVM parameters determine the model's performance. Consequently, SVM has been selected as a complementary model for predicting water quality.

Initial trials eliminate missing values. Remove missing values and apply machine learning. Table 3 demonstrates machine learning outcomes from deleting values that are missing from the dataset.

As per the findings, SVM achieves the greatest score for accuracy of 98% out of all the individual models. A 95% precision score, 99% recall score, 95% sensitivity score, and 99% F1 score are all attained via SVM. At 76% accuracy, 45% precision, 68% recall, 56% sensitivity, and 50% F1 score, BPNN is the worst performer. RBFNN attains 76% accuracy, 73% precision, 74% recall, 75% sensitivity, and 75% F1 in its computations. The results obtained by KNN are 93%, 92%, 95%, and 91% for precision, recall, accuracy, sensitivity, and F1 scores. Machine learning models that eliminate missing values produce poor results.Figure 6 illustrates machine learning model results after missing-value data elimination. The data clearly show that SVM outperforms the other models. We utilized KNN imputer. Data preparation finds missing values. The KNN imputer imputed missing data using Euclidean distance and mean for supplied values. Table 8 illustrates the results of the KNN imputer's data-tested machine learning model.

Table 8: Machine learning algorithm assessments with KNN imputer data

Model	Accuracy	precision	Recall	sensitivity	F1-score
RBFNN	0.76	0.73	0.74	0.75	0.75
BPNN	0.64	0.45	0.68	0.56	0.50
SVM	0.98	0.95	0.99	0.95	0.99
KNN	0.93	0.92	0.95	0.93	0.91

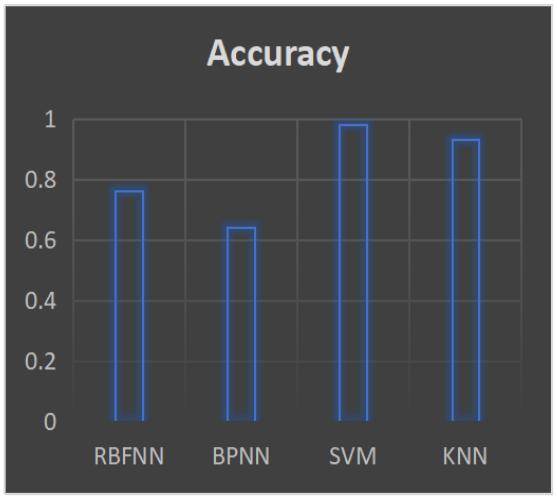

Fig 6(a)

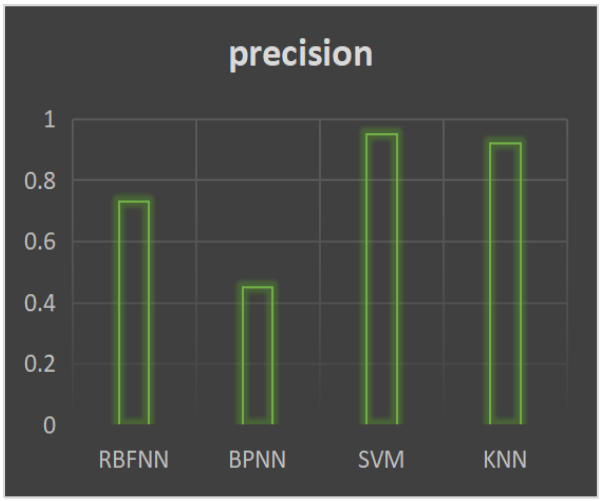

Fig 6(b)

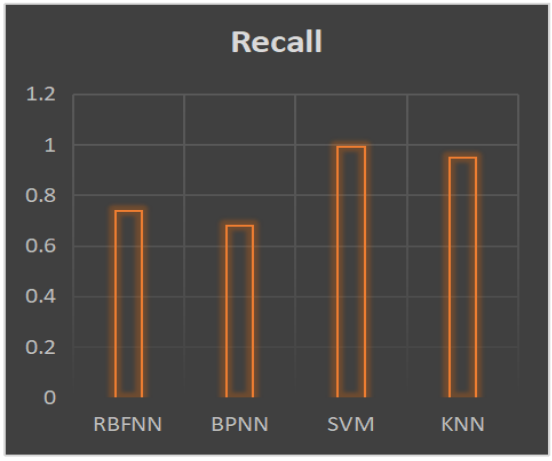

Fig 6(c)

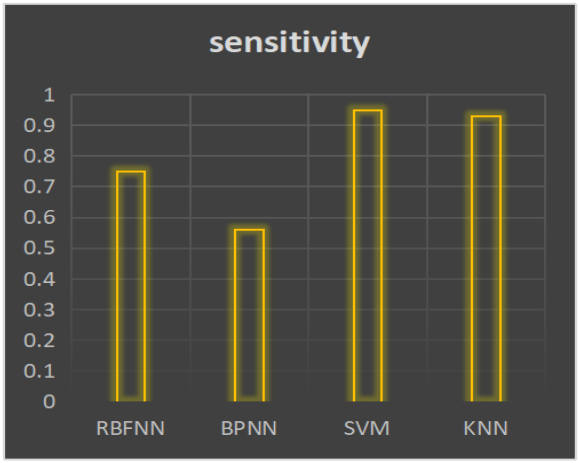

Fig 6(d)

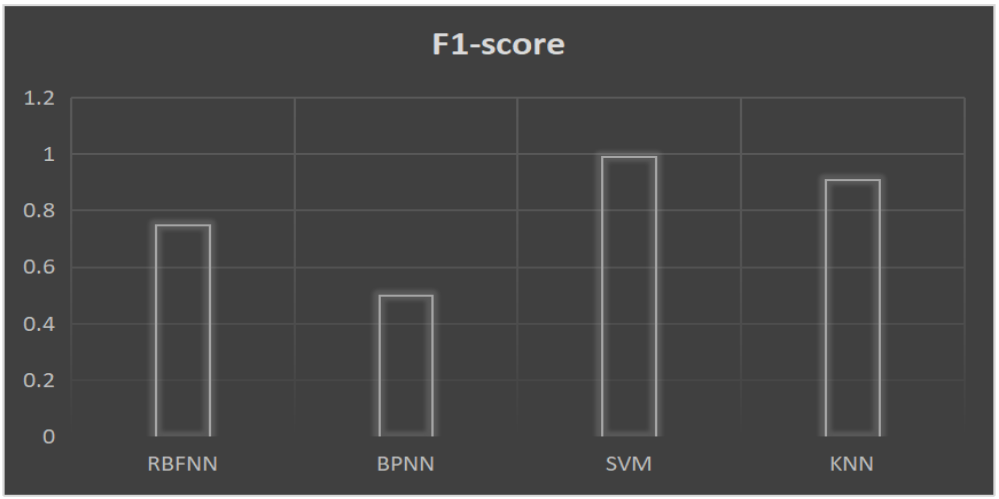

Fig 6(e)

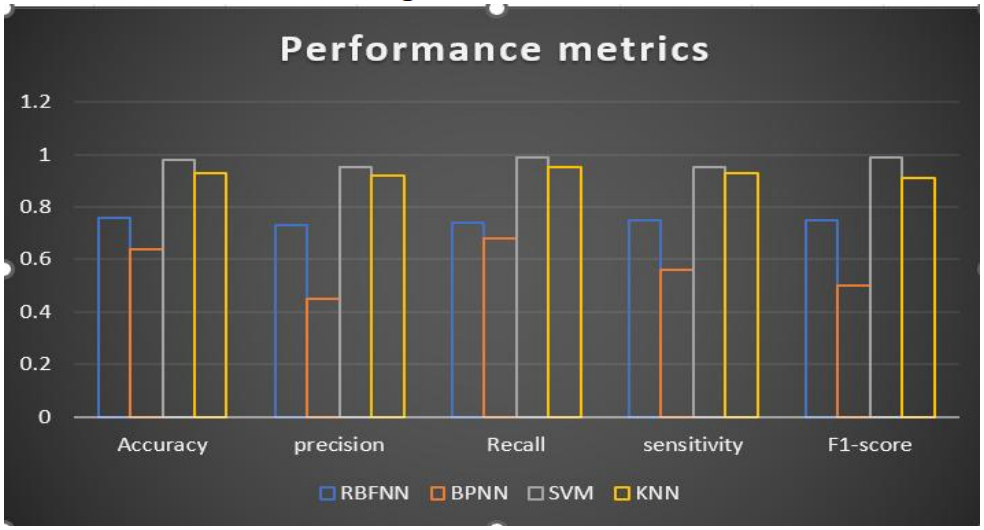

Fig 6(f): Model results from KNN imputer-filled datasets

KNN Imputer and also Data Missing Elimination Accuracy vs. All Learning Models. This section compares learning models with and without KNN imputers to provide a complete and understandable performance evaluation. Experimental results reveal that KNN imputers fill in missing data to improve learning models. These results are in line with models of learning that did not use the KNN imputer. The accuracy of every model of learning is compared in Table 9 using the KNN imputer, which removes missing data.

Table 9: Machine learning model accuracy with and without KNN imputer

Model	Accuracy	
	KNN imputer	Missing value deletion
RBFNN	0.85	79
BPNN	0.73	72
SVM	0.98	85
KNN	0.95	91

As illustrated in Figure 7, the KNN imputed dataset and machine learning method results after missing value removal are shown. The KNN imputer improves each independent learning model and the overall performance.

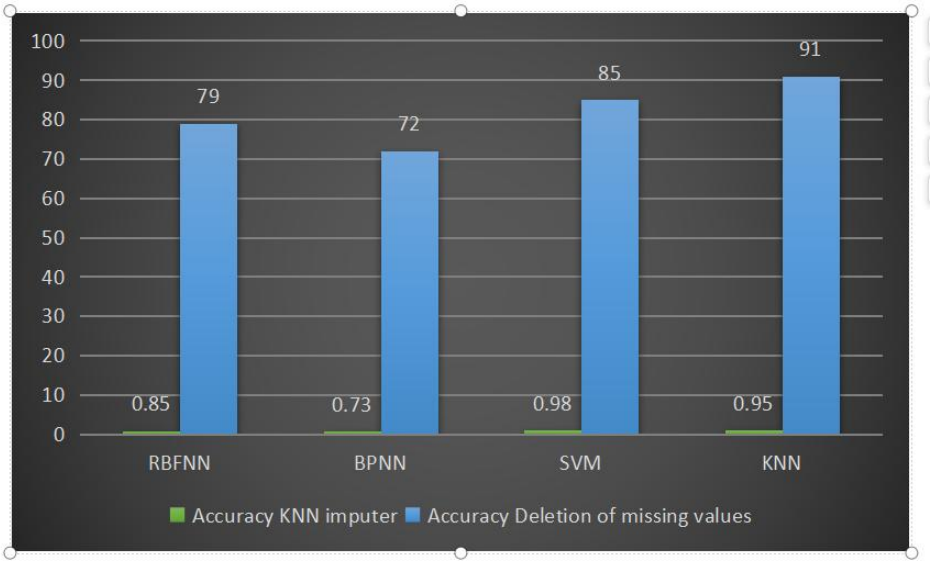

Fig 7: Graph of KNN imputer model performance

For the published data, SVM demonstrated outstanding prediction capabilities. SVM was chosen to forecast an aquaculture body's water quality data in a commercial system. Figure 3 displays the forecast's outcomes, whereas Figure 8 displays the performance metrics that were determined in Table 10 and Figure 8(b) show that the pH has the lowest MSE and R square values (0.003 and 0.95, respectively). DO MSE is 0.001, and its R-squared is 0.96. For NH3-N, the R-squared result is 0.93, while the MSE is 0.001. NO3-N has 0.003 MSE and 0.92 R-squared. R-squared is 0.98, and MSE is 0.001 for NO2-N. The findings demonstrated that SVM produced outstanding prediction performance in practical situations, as evidenced by its forecasting accuracy for measurements on industrial aquaculture water remained high. Industrial aquaculture systems can be monitored and prognosed using SVM prediction. We recommend short-term forecasting for better predictions.

The system was more reliant on DO to obtain pH predictions and was more sensitive to pH changes for DO parameter calculation, according to the sensitivity analysis. Additionally, when NH3-N was predicted, the algorithm became more sensitive to NO3-N, and when NO3-N was present, the algorithm used became more sensitive to NH3-N. The NO2-N input was predicted by the model with more sensitivity compared to the NO3-N input.

Table 10: ML models predict water quality.

Water quality parameter / ML Model	DO		pH		NH3-N		NO3-N		NO2-N	
	MSE	R2	MSE	R2	MSE	R2	MSE	R2	MSE	R2
BPNN	0.006	0.70	0.062	0.62	0.065	0.84	0.019	0.58	0.005	0.49
RBFNN	0.003	0.85	0.041	0.85	0.021	0.73	0.004	0.72	0.451	0.84
SVM	0.001	0.96	0.003	0.95	0.001	0.93	0.003	0.92	0.001	0.98
KNN	0.004	0.94	0.052	0.75	0.056	0.68	0.041	0.85	0.74	0.92

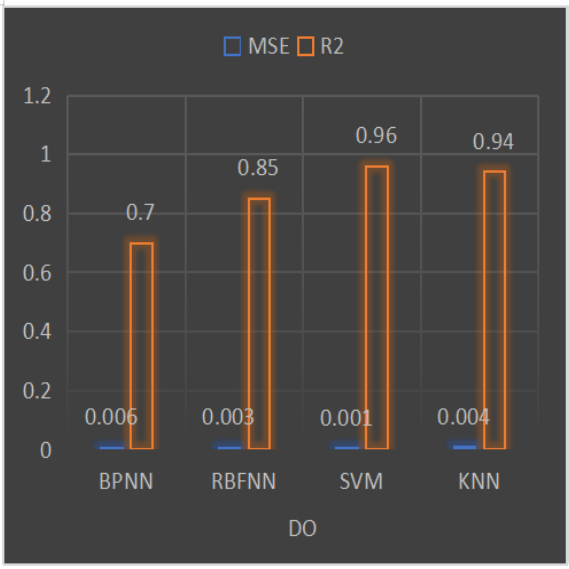

Fig 8 (a)

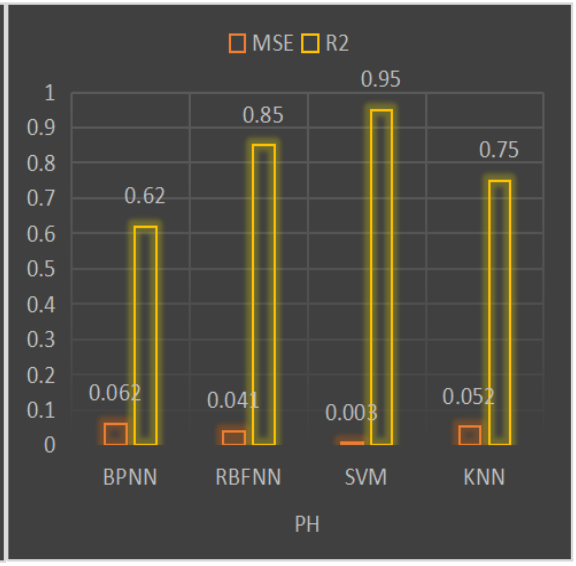

Fig 8 (b)

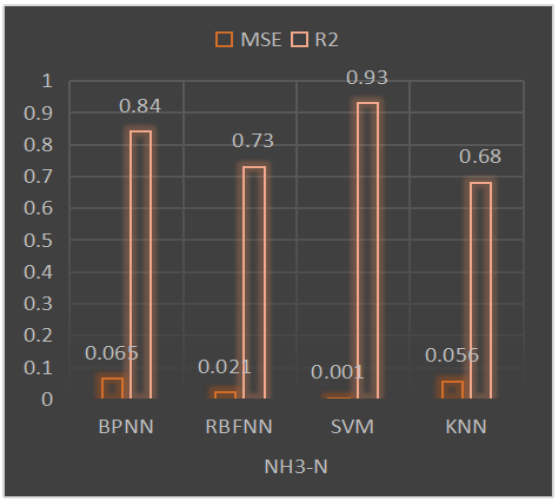

Fig 8 (c)

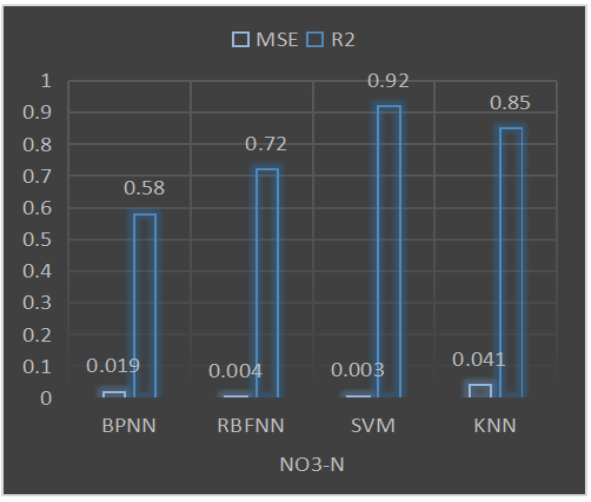

Fig 8 (d)

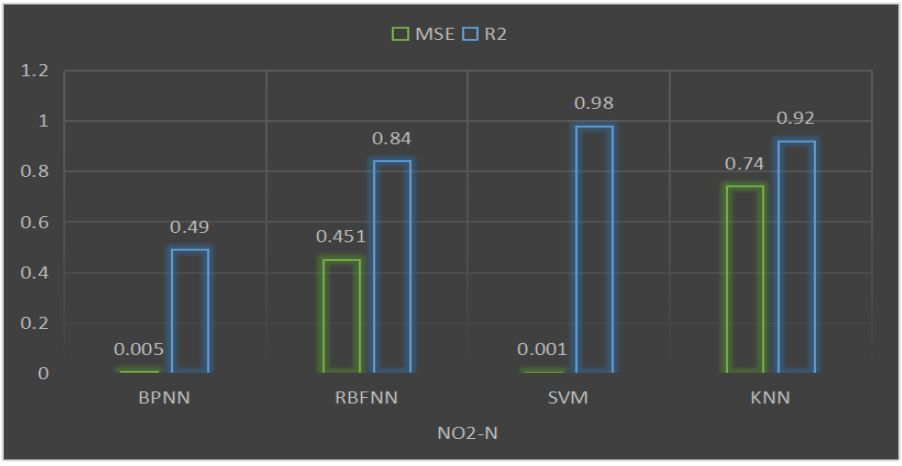

Fig 8 (e)

Fig 8: Performance indicators including MSE & R-Square of different algorithms

SVM beat all four prediction models in industrial aquaculture data. This study used SVM to improve prediction accuracy. For example, when predicting ground water levels and the health of the coastal waters. SVM with 98% accuracy is highly recommended for industrial aquaculture for both water

quality prediction and monitoring. More accurate and dependable prediction results can be obtained by using parameter optimisation.

For the data that was published, SVM demonstrated exceptional prediction capabilities. Therefore, in the industrial aquaculture system, SVM was chosen to forecast the aquaculture body's water quality data. The results of the forecast are displayed in Figure 9.

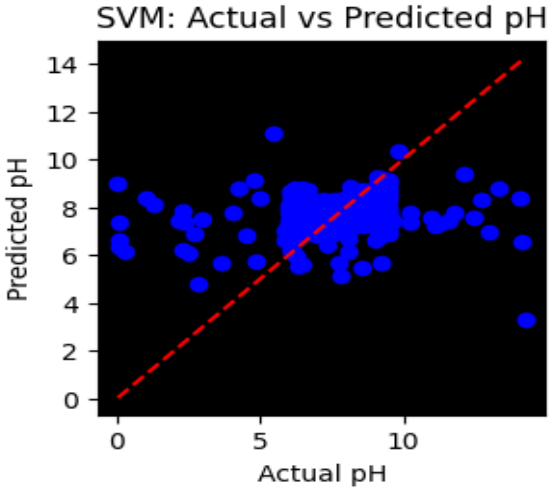

Fig 9 (a)

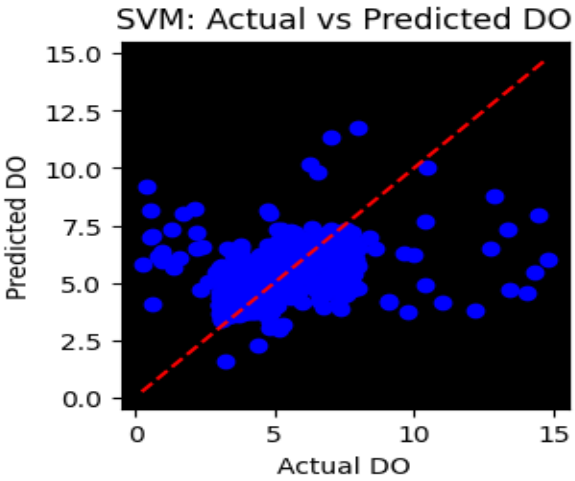

Fig 9 (b)

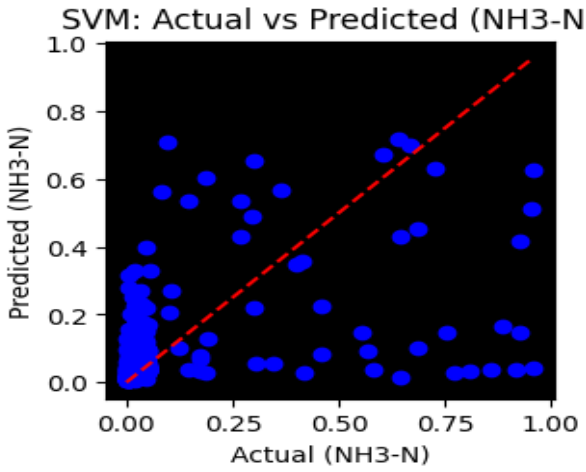

Fig 9 (c)

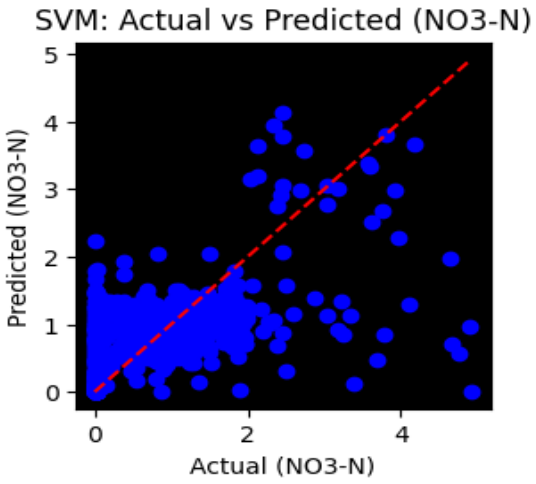

Fig 9 (d)

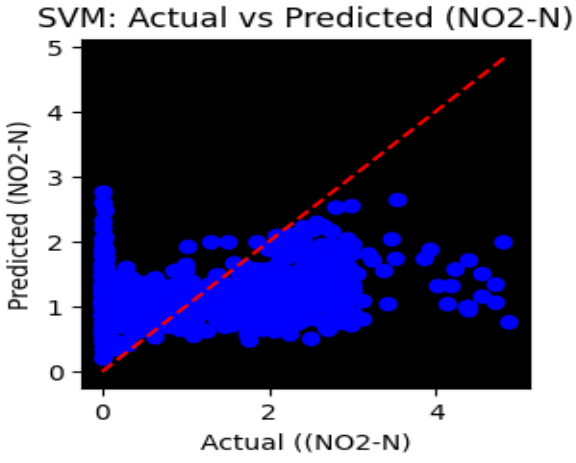

Fig 9 (e)

Fig 9: Simulation and prediction of SVM on water quality parameters. (a): pH ; (b): DO; (c): NH3 -N; (d): NO3 -N; (e): NO2 -N.

10. CONCLUSION

Industrial aquaculture requires clean water. Due to high expenses, many aquacultural systems cannot afford in situ or real-time surveillance. To regulate production, industrial aquaculture systems must predict water quality. This work simulated and predicted DO, pH, NH4-N, NO3-N in, and NO2-N using four machine-learning models: BPNN, RBFNN, SVM, and KNN. A KNN imputer handled values that were missing in the dataset. Both KNN imputer and missing data reduction are done during trials. The findings indicate that the KNN imputer is a superior option for filling in the values that are missing because removing the value that is missing data results in information loss that impacts the models' performance. According to the main conclusions, SVM performed better than the other models despite requiring less data and demonstrating more stability. Although RBFNN was comparatively good at predicting individual indicators, it was unstable and had an overwhelming accuracy gap. BPNN and KNN failed to predict water quality. SVM accurately predicted water quality.

11. FUTURE SCOPE

The future of ML in water quality prediction offers exciting possibilities for enhancing environmental management. Key advancements include real-time monitoring, where ML models can be integrated with sensors to provide immediate predictions and enable quick responses to water quality issues. This integration would help detect pollution or changes in water parameters like Dissolved Oxygen (DO) and pH as soon as they occur. Additionally, improvements in model accuracy and robustness will come from better handling of noisy, incomplete, or imbalanced data. Techniques like ensemble learning and deep learning can improve prediction reliability and efficiency. Another significant development is incorporating climate change effects into predictions, allowing models to forecast long-term changes in water quality due to shifts in temperature and weather patterns.

Furthermore, cross-regional adaptability will allow ML models to work effectively across different geographic locations, ensuring that predictions are relevant globally. Finally, linking ML models to decision support systems (DSS) will help automate responses based on predictions, improving water quality management and providing faster, data-driven decisions.

These advancements will make ML models more accurate, scalable, and practical for real-time, global water quality monitoring, improving the overall management of water resources.

References

1. Amirgaliyev, B., Kozbakova, A., Omarova, P., Merembayev, T. and Amirzhan, K.“Development and experimental study of an intelligent water quality monitoring system based on the internet of things.” Bulletin of Electrical Engineering and Informatics, 14(1), pp.761-773. 2025
2. Karumanchi, K.P., Mhatre, J., Lee, A. and Lee, “AI/ML-Based Water Quality Monitoring Mobile App for Predicting E. coli in Surface Waters.” In 2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM) (pp. 1-8). IEEE. 2025 January.
3. Shafi, J., Ijaz, R., Koul, A. and Ijaz, M.F. “Data-driven water quality prediction using hybrid machine learning approaches for sustainable development goal 6.” Environment, Development and Sustainability, pp.1-39. 2025.
4. Baena-Navarro, R., Carriazo-Regino, Y., Torres-Hoyos, F. and Pinedo-López. “Intelligent prediction and continuous monitoring of water quality in aquaculture: Integration of machine learning and Internet of Things for sustainable management.” Water, 17(1), p.82. 2025.
5. Shams, M.Y., Elshewey, A.M., El-Kenawy, E.S.M., Ibrahim, A., Talaat, F.M. and Tarek, Z. “Water quality prediction using machine learning models based on grid search method.” Multimedia Tools and Applications, 83(12), pp.35307-35334. 2024.
6. Joshi, T., Joshi, B., Syed, K., Vijayarani, S., Kumar, S. and Naval, N. “Recommendation System in Industrial data analysis using Machine Learning Techniques.” In 2024 IEEE North Karnataka Subsection Flagship International Conference (NKCon) (pp. 1-5). IEEE. 2024, September.
7. Kumar, B.V. and Rao, P.G.K. “An effective hybrid attention model for crop yield prediction using IoT-based three-phase prediction with an improved sailfish optimizer.” Expert Systems with Applications, 255, p.124740. 2024.
8. Chen, J., Li, H., Felix, M., Chen, Y. and Zheng, K. “Water quality prediction of artificial intelligence model: a case of Huaihe River Basin, China.” Environmental Science and Pollution Research, 31(10), pp.14610-14640. 2024.
9. Dewi, D.A., Wei, A.S., Lin, L.C. and Heng, C.D. “Water Quality Prediction using Random Forest Algorithm and Optimization.” Journal of Applied Data Sciences, 5(3), pp.1354-1362. 2024.
10. Pandya, H., Jaiswal, K. and Shah, M. “A comprehensive review of machine learning algorithms and its application in groundwater quality prediction.” Archives of Computational Methods in Engineering, 31(8), pp.4633-4654. 2024.

11. Chatterjee, T., Gogoi, U.R., Samanta, A., Chatterjee, A., Singh, M.K. and Pasupuleti, S., 2024. "Identifying the most discriminative parameter for water quality prediction using machine learning algorithms." Water, 16(3), p.481. 2024.
12. William, P., Oyebode, O.J., Ramu, G., Gupta, M., Bordoloi, D. and Shrivastava, A. "Artificial intelligence-based models to support water quality prediction using machine learning approach." In 2023 International Conference on Circuit Power and Computing Technologies (ICCPCT) (pp. 1496-1501). IEEE. 2023 August.
13. Grekov, A.N., Kabanov, A.A., Vyshkvarkova, E.V. and Trusevich, V.V. "Anomaly detection in biological early warning systems using unsupervised machine learning". Sensors, 2023 23(5), p.2687.
14. Karras, A., Karras, C., Sioutas, S., Makris, C., Katselis, G., Hatzilygeroudis, I., Theodorou, J.A. and Tsolis, D.,."An integrated GIS-based reinforcement learning approach for efficient prediction of disease transmission in aquaculture". Information, 202314(11), p.583.
15. Rizal, N.N.M., Hayder, G. and Yussof, S.,"River water quality prediction and analysis–deep learning predictive model's approach". In Sustainability challenges and delivering practical engineering solutions: resources, materials, energy, and buildings 2023 (pp. 25-29). Cham: Springer International Publishing.
16. Panigrahi, N., Patro, S.G.K., Kumar, R., Omar, M., Ngan, T.T., Giang, N.L., Thu, B.T. and Thang, N.T.,."Groundwater quality analysis and drinkability prediction using artificial intelligence". Earth Science Informatics, 202316(2), pp.1701-1725.
17. Swetha, P., Rasheed, A.H.K. and Harigovindan, V.P., "Random Forest Regression based water quality prediction for smart aquaculture". In 2023 4th International Conference on Computing and Communication Systems2023March. (I3CS) (pp. 1-5). IEEE.
18. Ghosh, H., Tusher, M.A., Rahat, I.S., Khasim, S. and Mohanty, S.N.,."Water quality assessment through predictive machine learning". In International Conference on Intelligent Computing and Networking 2023, February (pp. 77-88). Singapore: Springer Nature Singapore.
19. Prasad, D.V.V., Venkataramana, L.Y., Kumar, P.S., Prasannamedha, G., Harshana, S., Srividya, S.J., Harrinei, K. and Indraganti, S."Analysis and prediction of water quality using deep learning and auto deep learning techniques". Science of the Total Environment, 2022821, p.153311.
20. Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A. and Al-Shamma'a, A."Water quality classification using machine learning algorithms". Journal of Water Process Engineering, 202248, p.102920.
21. Ewuzie, U., Bolade, O.P. and Egbedina, A.O.,."Application of deep learning and machine learning methods in water quality modeling and prediction: a review". Current trends and advances in computer-aided intelligent environmental data engineering, 2022 pp.185-218.
22. Mokhtar, A., Elbeltagi, A., Gyasi-Agyei, Y., Al-Ansari, N. and Abdel-Fattah, M.K.,."Prediction of irrigation water quality indices based on machine learning and regression models". Applied Water Science, 202212(4), p.76.
23. Trach, R., Trach, Y., Kiersnowska, A., Markiewicz, A., Lendo-Siwicka, M. and Rusakov, K.," A study of assessment and prediction of water quality index using fuzzy logic and ANN models". Sustainability, 202214(9), p.5656.
24. Juna, A., Umer, M., Sadiq, S., Karamti, H., Eshmawi, A.A., Mohamed, A. and Ashraf, I.. "Water quality prediction using KNN imputer and multilayer perceptron". Water 2022, 14, 2592 [online]
25. Khan, M.S.I., Islam, N., Uddin, J., Islam, S. and Nasir, M.K.,."Water quality prediction and classification based on principal component regression and gradient boosting classifier approach". Journal of King Saud University-Computer and Information Sciences, 202234(8), pp.4773-4781.
26. Hmoud Al-Adhaileh, M. and WaselallahAlsaade, F.,."Modelling and prediction of water quality by using artificial intelligence". Sustainability, 202113(8), p.4259.
27. Tiyasha, Tung, T.M. and Yaseen, Z.M.,."Deep learning for prediction of water quality index classification: tropical catchment environmental assessment". Natural Resources Research,2021 30(6), pp.4235-4254.
28. Nordin, N.F.C., Mohd, N.S., Koting, S., Ismail, Z., Sherif, M. and El-Shafie, "A Groundwater quality forecasting modelling using artificial intelligence": A review. Groundwater for Sustainable Development,2021 14, p.100643.
29. Song, C., Yao, L., Hua, C. and Ni, Q.,."A water quality prediction model based on variational mode decomposition and the least squares support vector machine optimized by the sparrow search algorithm (VMD-SSA-LSSVM) of the Yangtze River, China". Environmental monitoring and assessment, 2021193(6), p.363.
30. Kulisz, M., Kujawska, J., Przysucha, B. and Cel, W.,."Forecasting water quality index in groundwater using artificial neural network". Energies, 202114(18), p.5875.