

## Machine learning-driven intelligent water quality assessment for enhanced drinking safety and real-time consumer awareness

Karthick Raghunath K. M.<sup>a</sup>, Surbhi Bhatia Khan<sup>b,c,d</sup>, Priya Govindarajan<sup>e</sup>, Mahesh T. R. <sup>a,\*</sup>, Mohammad Alojail<sup>f</sup> and Thippa Reddy Gadekallu<sup>g,h,i</sup>

<sup>a</sup> Department of Computer Science and Engineering, JAIN (Deemed-to-be University), Bengaluru 562112, India

<sup>b</sup> Department of Data Science, University of Salford, Manchester, United Kingdom

<sup>c</sup> University Centre for Research and Development, Chandigarh University, Mohali, Punjab

<sup>d</sup> Centre for Research Impact and Outcome, Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, 140401, Punjab, India

<sup>e</sup> Department of Computer Science, Amrita School of Computing, Amrita Vishwa Vidyapeetham, Mysuru Campus, Karnataka, India

<sup>f</sup> Management Information System Department, College of Business Administration, King Saud University, Saudi Arabia

<sup>g</sup> College of Mathematics and Computer Science, Zhejiang A&F University, Hangzhou 311300, China

<sup>h</sup> Division of Research and Development, Lovely Professional University, Phagwara, India

<sup>i</sup> Center of Research Impact and Outcome, Chitkara University, Rajpura, 140401, Punjab

\*Corresponding author. E-mail: t.mahesh@jainuniversity.ac.in

 MTR, 0000-0002-5589-8992

### ABSTRACT

As to the sphere of smart water management and managing water Internet of Things (IoT) systems, water condition safety for drinking is very important. The proposed methodology, known as the Smart Water Consumption Monitoring System (SWCMS), is based on the WaterNet dataset acquired from a standard data repository for training the selected machine learning (ML) models. For water quality parameters such as temperature, turbidity, pH, and some chemical concentrations, the system uses real-time sensors. At the testing phase, information received from the sensors is time-stamped, and with the utilization of applicable ML approaches, potential challenges; assessment of water quality is processed. This encompasses the employment of advanced instruments for the detection of water quality with concentration on pH and other chemical values through a detection accuracy rate of over 95% on any other signs of abnormalities. This processed information is further availed with the timestamps to the consumers' mobile phones through a user interface application for real-time awareness and timely response. With the aid of timely information about their drinking water, the SWCMS increases the water safety parameter by 90% and the overall consumer awareness by 92.5%, thereby creating an effective health parameter among the public.

**Key words:** cascade generalization, detection, machine learning, monitoring, sensor data, water quality

### HIGHLIGHTS

- The SWCMS uses sensor data and ML models to monitor sump water tanks, assessing key attributes like temperature, turbidity, pH, and chemical content.
- This proactive monitoring enhances public health and safeguards communities from potential hazards.
- The SWCMS aims to deliver critical insights into drinking water quality and potential pathogen presence for consumers.

## 1. INTRODUCTION

In the recent research arena, the field of smart water management has gained much attention and progress thanks to the incorporation of Internet of Things (IoT) (Mishra 2023) applications and machine learning (ML) concepts. The advancements made here have led to better means and ways of using technology in monitoring, quality, distribution, and consumption of water. The scarcity and pollution of freshwater bodies across the world called for the design and implementation of intelligent systems that would enhance the use of water to the maximum safety level (Nasir *et al.* 2022). Smart devices, such as those based on IoT, can proactively monitor and detect different parametric qualities of drinking water in real time. On the other hand, there are ML algorithms that can detect possible issues with water quality and predict possible

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

future trends (Lowe *et al.* 2022). This has, therefore, made IoT/ML a potent tool in management and solving water issues, making it an important field of study.

However, smart water management has so far not benefited from enough investigative work revealing the best ways of inspecting and evaluating the quality of drinking water, especially at home. There are many traditional water monitoring systems that are challenged due to the following reasons: the systems do not offer actual contextual data, real-time analysis, and appropriate feedback to consumers (Xu *et al.* 2022). These systems normally involve testing performed at some intervals, and this can be costly, time-consuming, and may involve some errors from the testers. However, the existing solutions do not align smoothly with the consumer applications as they were marketed and implemented mostly in business contexts, thus reducing their effectiveness. This indicates the need for solutions that integrate real-time data acquisition, robust data processing, visualization tools, and interfaces to produce meaningful information on water quality and usage for consumers (Talukdar *et al.* 2023).

To fill the said research gap, this study develops another methodology named the Smart Water Consumption Monitoring System (SWCMS), which has the potential to monitor sump water tanks and review the current condition of drinking water through the use of smart sensors and ML algorithms (Ajayi *et al.* 2022a). Using the WaterNet data set from IEEE Dataport (Ajayi *et al.* 2022b; Silva 2022), the SWCMS uses a multi-layered cascade generalization (CG) to improve the detection accuracy and conditional dependability of water quality predictions. The system needs to pass through two levels of cascading architecture, which include long short-term memory (LSTM), convolutional neural networks (CNN), and random forest (RF) in the first level. Next, such basic-level model outputs are combined and passed to the second-tier models like gradient boosting machines (GBMs) and stacked denoising autoencoders (SDAE) to enhance the detection. The last decisions of the model, together with timestamps, are sent to the consumer's mobile phone, and hence, consumers gain real-time awareness about the environment, which in turn can prompt action.

The rationale for this study arises from the imperative of safeguarding residents' supply of safe water in homes since it can be very risky because water quality in homes can be very disparaging as a result of such factors as degraded infrastructure, contamination, and lack of monitoring (Manjakkal *et al.* 2021). Thus, the core investigation of this study is to create an intelligent system that can thereby monitor water quality, learn from its results, and instantaneously give feedback to the consumer to make the proper decisions regarding water use. New SWCMS development, including design, implementation, and assessment of the effectiveness of the system, will aim to promote water safety, raise public awareness, and consequently decrease the number of water-borne diseases. This approach guarantees the system's comprehensiveness and, at the same time, provides its users with the most convenient interface by utilizing advanced sensors, ML models, and mobile applications.

Thus, the importance of this study is related to its capacity to change the approach to detect the conditions of the quality of drinking water. Regarding concerns within the scope of smart water management, the SWCMS fulfils a number of significant tasks due to real-time data analysis for water quality assessment. First, it painted the picture of a solution that can easily be implemented in different dwellings and maintain a reliable survey of water quality. Second, the incorporation of sophisticated ML methods improves the reliability of the system and increases the efficiency of the detection of anomalies' presence and the prognosis of future trends. Third, the utilization of an awareness mechanism in real-time helps the consumer to be informed on any given incident and take necessary action to avoid health complications. In conclusion, this study makes a modest contribution toward the objective of practicing sound and secure water use throughout the world based on prior research due to the water crisis and related issues that have been identified as significant in the general population's health and the environment.

The following strategic goals of the SWCMS concept have been identified as the core objectives of the system:

- The SWCMS, with effective sensor data and efficient ML models, is intended to continuously oversee sump water tanks and detect/evaluate/record important attributes of the water, such as temperature, turbidity, pH, and other vital chemical content. The system's desire to achieve a detection accuracy rate of more than 95% allows for the immediate detection of any chemical pathogens in water quality. This preemptive monitoring significantly improves the health of communities and protects the population from potential threats.
- The SWCMS intends to provide consumers with vital information concerning the quality of drinking water and the presence of potential pathogens. In the context of the analyzed framework, such data are time-stamped and made available to consumers for real-time informational awareness. In this way, the ultimate goal of the system is to enhance consumer

awareness and also water safety parameters to the maximum extent, thereby raising the health parameters among the public.

However, the proposed methodology is evaluated with the utilization of a well-defined ‘WaterNet’ dataset from IEEE DataPort (Ajayi *et al.* 2022b), which can help future research to build upon the present work. In a real-time testbed, it is imperative to introduce certain routines, such as checking and calibrating sensors frequently, to produce accurate results of water quality. Calibration eliminates variability that is often caused by the drift of the sensor, while maintenance removes variability that emanates from fouling, impacting the performance of the instrument to produce accurate data at different time intervals. Last but not least, the research offers an extensive assessment of the SWCMS and references the effectiveness and relevance of the model to water safety and relevant populations’ health concerns.

The paper is formatted as follows. Commencing with a comprehensive introduction of the paper, the discussion is followed more specifically by a set of reviews of existing work to build the framework for conducting research. Next, the section on methodology describes the dataset that has been utilized, the data pre-processing process, and the CG process in which it has been incorporated, including Layer 1, comprising lower-level models; intermediate fusion, which is Layer 2 and Layer 3 concerning higher-level models. The performance evaluation and discussions come next, during which the findings are scrutinized and expounded. Last, the paper discusses the study’s implications as a precise conclusion and concerns for further research.

## 2. RELATED WORK

Manjakkal *et al.* (2021) and Lakshmikantha *et al.* (2021) used low-cost sensors, along with smart algorithms for structural health monitoring (SHM) of pipes using standardized data analysis and prediction. This strategy entails the improvement of the data scope and comparability across the strata of all water bodies across the world. However, some limitations are the absence of any clear protocol for data collection, spatial-temporal characteristics, and the need for enhanced and denser sensor systems. These challenges call for a constant enhancement of strategies in sensor placement and information processing in order to achieve effective water quality monitoring (WQM). Lakshmikantha *et al.* (2021) and Garrido-Momparler & Peris (2022) described IoT-based synchronous water quality meandering at low cost. The base approach of the core methodology involves using sensors to quantify the pH, turbidity, conductivity, and temperature of water. These sensors are interfaced with an Arduino microcontroller that forwards the measured information to a cloud server for constant evaluation. Hence, the study aims to offer a cheap and effective method for the achievement of undesirable water safety. Nonetheless, some of its limitations include the constant dependence on Internet connection throughout the analysis and data collection process, and possible difficulties in calibrating and/or maintaining the sensors that are used during the measurements. Garrido-Momparler & Peris (2022) and El-Shafeiy *et al.* (2023) concentrated on the connectivity of smart sensors within the IoT infra and cloud computing for monitoring the concerned subjective conditions, especially water quality. The fundamental method used for managing these applications consists of using smart sensors with portability and low-power consumption that have the ability to independently transfer information to cloud processing systems and analytics for higher-level evaluation. This will help improve the frequency and accuracy of environmental monitoring. However, some of the major drawbacks are associated with power consumption, price, and common problems that arise in interactions between different systems, which can be solved only by establishing the generally accepted rules of their functioning and corresponding protocols of communication. To identify outliers in water quality data, El-Shafeiy *et al.* (2023) and Syrmos *et al.* (2023) put forward multiple convolutional networks (MCN)–LSTM; a complex approach that uses both MCNs and LSTM. The core methodology is based on deep learning, which implies the analysis of time series acquired by a network of IoT-based sensors that consider both spatial and temporal characteristics. The major concern here is to ensure that early and accurate signals of fluctuating or negative data are checked for water quality. However, a major disadvantage of this method is that it is computationally intensive and requires large amounts of data for testing, which at times is not feasible. More work is recommended to fine-tune these requirements and to increase practical relevance. Syrmos *et al.* (2023) and Chinnappan *et al.* (2023) developed a systematic approach for an efficient IoT architecture used for real-time water monitoring via long-range wide-area network (LoRaWAN) with the help of the ML algorithm. The core technique includes employing flow meters and different water quality sensors to measure consumption rates and quality features, wireless data transfer by LoRaWAN, and employing cloud computing and different ML predictive models. This system is intended to strengthen water management altogether, both in the urban and rural areas, by offering information concerning the use of water and quality

in real time. A limitation of this approach is that the sensors have to be calibrated from time to time, and their maintenance is also an act that may add to the operation costs and complicate the system. Chinnappan *et al.* (2023) and Shahra & Wu (2023) proposed an IoT architectural model for real-time chlorine level checks and forecasts in water sources. The basic working style consists of measuring the values of temperature, flow, and chlorine level using sensors, processing them through Raspberry Pi microcontroller, sending the information to the cloud, and applying the fuzzy logic-based decision tree (FL-DT) technique for prediction. Its goal is to maintain a constant supply of safe drinking water by regularly checking the amount of chlorine. However, they rely on pure data quality and availability, which can be a constraint, and the system requires a lot of computational power. However, the work should be directed to improve these aspects to increase the practicability and expansibility of the framework. Shahra & Wu (2023) and Rana *et al.* (2023) developed a dynamic methodological framework for solving the optimal sensor placement problem for near-actual monitoring of water contamination in the Water Distribution System. At the core of the plan, an evolutionary algorithm (EA) is applied to identify the strategies for placing sensors with reference to different contamination situations, with objectives of shortest detection time and wider coverage. The study then proves the usefulness of such an approach through the use of case studies such as the Battle of the Water Sensor Network and a real-life case backdrop in Madrid. However, the limitation was observed in terms of the density of the EA and its computation, where substantial resources would be needed, and perhaps the algorithm would not be scalable for larger networks. Rana *et al.* (2023) and Jallow (2024) described a method using artificial neural networks (ANN) and LSTM models for water quality prediction and monitoring. These models are developed using seven critical constraints, namely dissolved oxygen, temperature, conductivity, pH, turbidity, total dissolved solids (TDS), and chlorides. The emphasis is put on the appropriate classification and prognosis of the water quality index (WQI), with prominent rates of success shown via measures such as the mean absolute error (MAE) and mean squared error (MSE). Nonetheless, some disadvantages of this study are high computational complexity, the need for extensive data for training and testing the models, and the need for further optimization and application to real-life case scenarios. Jallow (2024) and Ruiz-Moreno *et al.* (2023) introduced a new concept called learning from observations to improve water efficiency and life-artificial intelligence, which has utilized ML strategies along with microscopic cameras and bioluminescent sensors with the intent of improving a way of assessing water quality. The main approach involves sampling water, using sensors to determine adenosine triphosphate (ATP) and nitrate, as well as using TensorFlow Javascript to devise the ML models. According to this theoretical framework, ATP levels and bacterial concentration are expected to be strongly related exclusively. However, the dimensional approach of the study is yet in the conceptual stage and does not present real-world application, which is a weakness of the study.

### 3. METHODOLOGY

In the real-time scenario, the data are processed at the SWCMS, which receives constant updates on the water quality measurements from different sensors. These sensors send their data to a certain central processing unit (CPU) through established communication protocols such as the cellular data network or through Wi-Fi. The data are cleaned in order to remove noise and outliers before feeding it into CG architectural models (Seu *et al.* 2022) for further processing. The sampled results are automatically transmitted to a cloud server, which in turn makes the results available through the use of a mobile application in the form of notifications and a comprehensive water quality analysis. This real-time processing helps in having timely awareness and quick action to sustain the standard of safe drinking water. For empirical purposes, a well-defined dataset is employed to evaluate the models' performance.

Figure 1 illustrates the architecture of the proposed SWCMS. From the architecture, the CA uses systematic decision-making processes in which lower-level models' forecasts become the inputs to higher-level models, which strengthens the SWCMS architecture. Overall, the technique in this study analyses raw sensor data to obtain spatial features through CNNs and temporal features through LSTMs and performs RF to extract useful features from them. The features generated from these models are merged together to make the final convention called a feature set, which is used for the next-level models.

Then, the other higher-level models include GBMs and SDAE, which build on the previous models to add more value to the intra-dataset interaction and accuracy of the detections. The final output is the aggregation of these higher-level models, which gives a comprehensive detection to the customers to create real-time awareness and action about the quality of water through a mobile application. At the same time, each layer uses the strengths of individual models at different stages, and the application of this approach creates a reliable and accurate system for monitoring and evaluating drinking water conditions.

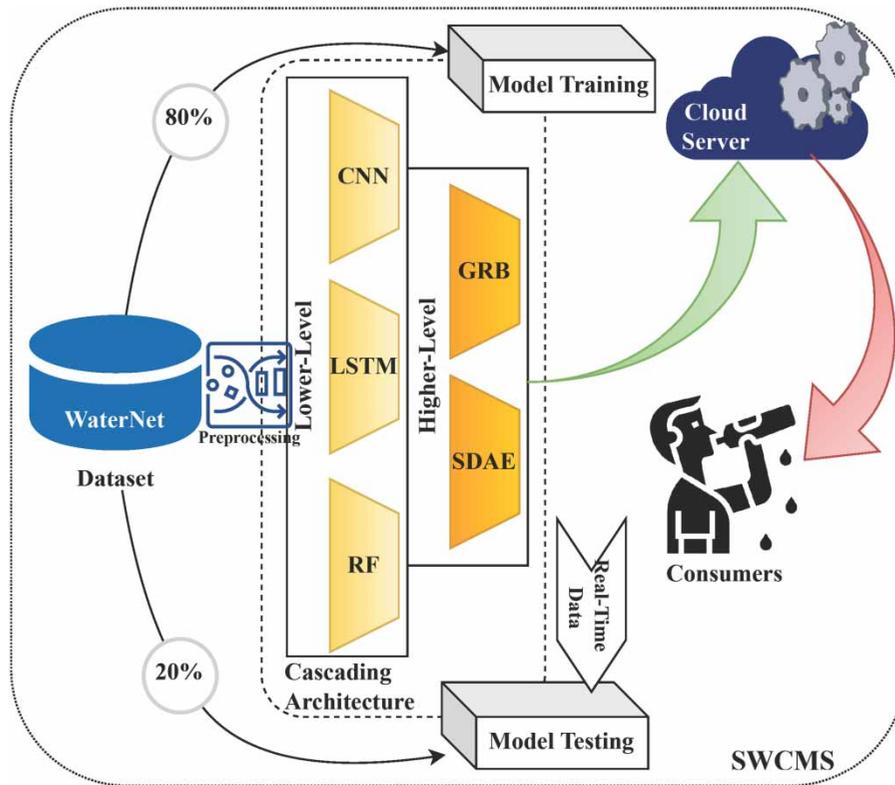


Figure 1 | Conceptual architecture of SWCMS.

Table 1 | Specifications and vital parameters of sensors utilized in the testing phase

Sensor		Range	Accuracy	Resolution
pH sensor (in pH)		0–14	± 0.01	0.01
Selective electrode (in mg/L)	Sodium ion	0.1–10,000	± 2% of reading	0.1
	Magnesium ion	0.1–1,000		
	Calcium ion			
	Chloride ion			
	Potassium ion			
	Carbonate ion			
	Sulfate			
TDS meter (in mg/L)		0–5,000		1
EC meter (in µS/cm)		0–2,000	± 1% of reading	1
Hardness sensor (in mg/L)		0–500	± 2% of reading	1
Temperatures (°C)		– 10 to 100 °C	± 0.1 °C	0.1 °C
Multi-parametric water quality meter	EC	0–2,000	EC ± 1%	EC 1
	TDS	0–5,000	TDS ± 2%	TDS 1
	pH	0–14	pH ± 0.01	pH 0.01
	Temp	10–100	Temp ± 0.1	Temp 0.1

In addition, Table 1 depicts the summary of the sensors and the reading adopted in the phase of SWCMS testing, with details concerning range, accuracy, resolution, and the unit of measure. These specifications are important to achieve the needed accuracy of measuring parameters of water quality.

**Table 2** | Prominent training parameters of the WaterNet dataset

Attribute	Specification
ID	Unique identifier per water sample
pH	Indicates the alkalinity or acidity of the water (6.5–8.5 for drinking water)
Sodium (Concentrated as mg/L)	Sodium ions
Magnesium	Magnesium ions
Calcium	Calcium ions
Chloride	Chloride ions
Potassium	Potassium ions
Carbonate	Carbonate ions
Sulphate	Sulfate ions
TDS	Inorganic and organic components in water (combined content)
EC ( $\text{\AA}\mu\text{S/cm}$ )	Water's ability to conduct electricity (related to dissolved salts)
Total hardness (TH)(mg/L)	Concentration of magnesium and calcium salts
WQI	Composite index representing overall water quality
Portability	Flag: 1 (potable), 0 (not potable)

### 3.1. Dataset

The dataset employed to work with the SWCMS is 'WaterNet', which is obtained from IEEE Dataport (Ajayi *et al.* 2022b). This dataset consists of 718 records and contains 14 essential features of water quality samples that dictate the potability of water. Through the application of all these attributes, systematically enhanced through a process known as CG, the SWCMS can deliver a timely and reliable assessment of drinking water quality, hence guaranteeing the consumers' safety and information. Table 2 represents the vital parameters, which are considered for training purposes.

The selection of sensors specified in Table 1 is associated with water quality characteristics, including pH, sodium, magnesium, and electrical conductivity (EC), which point to the safety and potability of water. These sensors help improve the accuracy of the overall system because each attribute is monitored in real-time if the SWCMS is to detect and combat possible water quality problems effectively. The current and steady checking of a broad scope of parameters makes it possible to assess the WQI and portability and maintain the potential to acquire organic and beneficial consequences for consumers.

### 3.2. Pre-processing

Analyzing and preparing the data for the SWCMS requires some important stages to provide precise and viable results in the ML models. First of all, the data cleaning is conducted, during which missing values are addressed with mean imputation (Singh & Singh 2022), replacing them with the mean of the given attribute. Outliers are found using the Z-score method (Huang *et al.* 2022) with data points that have a Z-score of more than three removed or corrected. The next operation is data normalization, which must bring the data within the range of 0–1 through min-max scaling.

This technique ensures that all attributes like sodium, pH, magnesium, and others are measured on the same scale, which is essential for CG. The dataset is then divided into training, validation, and test data sets so as to have valid model estimations. Feature engineering is carried out to make new features relevant or to convert an existing one to help detect patterns from the dataset. This kind of pre-processing also has the advantage of ensuring that the input data is well pre-cleansed and normalized for the subsequent training of the suggested ML models.

### 3.3. Cascade generalization

CG is the process of building a hierarchy that includes several models, where the prediction of models at a lower level is used by models at a higher level. In the current study, CG enhances the use of the three core layers. These are the lower-level models, the intermediate fusion, and the higher-level models. Layer 1 incorporates methods such as CNN, LSTM, and RF.

### 3.3.1. Layer 1 (lower-level models)

**CNN:** Here, the application of CNNs proves their effectiveness because raw sensor data undergoes rapid processing while important spatial patterns of the water quality parameters are identified. It proves that they are efficient in capturing local patterns and co-relations of the data in those regions. This raw data are fed into the convolutional layers, where the filter feature maps are used to detect features like trends in pH, sodium, and other chemical specifications. The outputs from the convolutional layers are posted through pooling layers to down-sample and emphasize major features. Last of all, fully connected layers enable these features to be transformed into the next process format. The outcome of this process is a feature map that holds some of the spatial information from the sensor data, and such an outcome gives confidence to enhance the effectiveness of the proposed methodology. The primary computation of this process is expressed as a raw sensor data matrix,  $\mathbb{I}_{\text{sensor}} \in \mathbb{R}^{r \times c}$ , where  $r$  and  $c$  indicate the sample size and feature count, respectively. Thus, to extract the spatial features, the standard convolution operation (Borup *et al.* 2023) is applied as follows, and the corresponding feature map outcome is represented in the following equation:

$$\mathbb{I}_{\text{cnn}} = f_{\text{cnn}}(\mathbb{I}_{\text{sensor}}) \quad (1)$$

$$\mathbb{I}_{\text{cnn}} \in \mathbb{R}^{r \times k_{\text{cnn}}} \quad (2)$$

**LSTM:** In the case of the temporal relation among various sensor values, recurrent neural networks (RNNs) are used, specifically LSTM networks. It is most applicable to cases of time series data since knowing the previous observations is a critical component of LSTMs. The raw sensor data  $\mathbb{I}_{\text{sensor}} \in \mathbb{R}^{r \times c}$  is passed into the LSTM layers, which retain some earlier inputs to establish and follow changes in the data pattern. This makes the model gain comprehension of how specific water quality parameters evolve over time, which is crucial in achieving desired results in terms of anomaly detection and/or prediction. The final output of the model  $\mathbb{I}_{\text{lstm}} \in \mathbb{R}^{r \times k_{\text{lstm}}}$  is in the form of temporal features that capture the sequence inherent in the sensed data. The major operational process (Borup *et al.* 2023) of this model is represented in Equation (3).

$$\mathbb{I}_{\text{lstm}} = f_{\text{lstm}}(\mathbb{I}_{\text{sensor}}) \quad (3)$$

**RF:** RF is incorporated in the lower-level models so as to perform feature selection as well as initiate predictions. RF is an extension of the bagging method where, during the construction of decision trees, an ensemble of them is built, and during classification, the mode of the classes is returned. It also has solutions for noisy data and explains which features are important in the dataset. Through the training of the RF model, thereby mapping the raw sensor data  $\mathbb{I}_{\text{sensor}} \in \mathbb{R}^{r \times c}$  to the target class, the prediction probabilities or feature importance for classification is obtained, thereby determining attributes of significance to the water quality,  $\mathbb{I}_{\text{rf}} \in \mathbb{R}^{r \times k_{\text{rf}}}$ . Thus, during the training phase, the feature selection and robust detection are carried out using the standard operational process (Manoharan *et al.* 2022) as depicted in the following equation:

$$\mathbb{I}_{\text{rf}} = f_{\text{rf}}(\mathbb{I}_{\text{sensor}}) \quad (4)$$

### 3.3.2. Intermediate fusion (layer 2)

As for the layers of the intermediate fusion layer, the output of the lower-level models is concatenated into a compound feature set. This entails the feature maps resulting from the CNNs, the temporal features from the LSTMs, and the probability of a prediction or the feature importance score if it was derived from the RFs. The fusion process lays the foundation of a highly comprehensive multiple features vector based on the strength of each lower-level model. It means that the present fused feature set includes all the water quality parameters, spatial and temporal distributions, and the most important features extracted by the RF model. These feature vectors are then prepared for feeding as input to the higher-level models using this comprehensive feature vector. Equation (5) expresses the output concatenation process of lower-level approaches, which tends to form a comprehensive feature set:

$$\mathbb{I}_{\text{fused}} = (\mathbb{I}_{\text{cnn}}, \mathbb{I}_{\text{lstm}}, \mathbb{I}_{\text{rf}}) \in \mathbb{R}^{r \times [k_{\text{cnn}} + k_{\text{lstm}} + k_{\text{rf}}]} \quad (5)$$

The intermediate fusion layer also improves the models' robustness and neutralizes the drawbacks of each lower-level model. For instance, CNNs are capable of learning localized spatial information, but they fail to learn the temporal patterns learned by LSTMs. Likewise, RFs aid in highlighting the most important features, suppressing noise, and preventing the model from over-focusing on irrelevant parts of the data provided. In this way, those various outputs are combined in the intermediate fusion layer to minimize over-reliance on specific patterns and increase robustness across a wide range of water quality cases. Furthermore, the fusion process allows the model to learn to adjust to changes in data occurrence (in an optimal way), thus reducing differences and fluctuations in sensor data, which is essential in real-time context as a result of unexpected noises or square data. The collective and continuous input that it involves makes the results obtained more accurate. Therefore, it is more reliable, thereby improving the capacities of the SWCMS in delivering timely and accurate water quality reports.

### 3.3.3. Layer 3 (higher-level models)

**GBM:** In all these higher-level approaches, GBMs (Dong *et al.* 2022) are then employed to improve the results from the fused feature set. GBMs are extremely effective ensemble learning mechanisms, which depend on building models incrementally, each new model trying to minimize the errors of the previous ones. Various types of features are fused together ( $\mathbb{I}_{\text{fused}}$ ), and then a GBM is trained on these fused features because the model can learn interactions between attributes for higher accuracy, which is highlighted as an operation expression via Equation (6). As a result of GBM, the model delivers a set of refined predictions ( $\hat{Y}_{\text{gbm}}$ ) that are enhanced by the iterative error-correction system, which constitutes boosting methods:

$$\hat{Y}_{\text{gbm}} = f_{\text{gbm}}(\mathbb{I}_{\text{fused}}), \quad \text{where } \hat{Y}_{\text{gbm}} \in \mathbb{R}^r \quad (6)$$

**SDAE:** SDAEs (Wang *et al.* 2022) are also used in higher-level models to improve the feature representations extracted from the intermediate fusion layer. SDAEs are generative restricted Boltzmann machines that attempt to learn a good feature representation by adding noise to the input and training the network to recognize the noiseless version of the data. This process aids the model in getting familiarized with mere features that improve its generality, and it works out this aspect. The SDAE then takes this fused feature set ( $\mathbb{I}_{\text{fused}}$ ) and processes it to gain a further denoised version of the data. The output from the SDAE is a set of enhanced features ( $\hat{Y}_{\text{sdac}}$ ) that provide deeper patterns on the input features referring to water quality parameters:

$$\hat{Y}_{\text{sdac}} = f_{\text{sdac}}(\mathbb{I}_{\text{fused}}), \quad \text{where } \hat{Y}_{\text{sdac}} \in \mathbb{R}^r \quad (7)$$

The final result is achieved by using the results from both the GBM and the SDAE proposed in this study. This is achieved through a simple summation in which the outputs of the two models are each weighted before being summed up in order to produce the final output. The weights ( $\vartheta$ ) are tuned through cross-validation so as to allow the network to perform at its best. Equation (8) exhibits that this combination utilizes the merits of both the GBM and SDAE; therefore, the accurate and reliable detection of water quality is enabled. It is then disseminated through the consumer's mobile application, keeping the user informed and enabling appropriate action concerning the water quality in real time:

$$\hat{Y} = \vartheta \hat{Y}_{\text{gbm}} + (1 - \vartheta) \hat{Y}_{\text{sdac}} \quad (8)$$

This CG mechanism also had the advantage of blending positive aspects of the various models in the SWCMS for training as well as the testing phases.

## 4. PERFORMANCE EVALUATION AND DISCUSSIONS

In order to assess the feasibility of the SWCMS model quantitatively, the software and the hardware specifications should be properly selected to obtain satisfactory results and the highest accuracy. The general software used is Python v3.9 with TensorFlow v2.8 for CNNs and LSTMs and Skikit-learn v0.24.2 for RF and GBM; PyTorch v1.11 is used for SDAE. Also, the Jupyter Notebook v6.4.3 is used in the interactive development and experimentation processes. Pandas v1.3.3 and NumPy v1.21.2 are used for data processing, while Matplotlib v3.4.3 is used for data visualization. The aspects that must be taken into consideration in terms of configurative hardware are high-performance computing with NVIDIA (Tesla V100-32 GB

for the GPU RAM), and a minimum of RAM of 64 GB with Intel Xeon processors with 16 cores for deep-learning model training. Features include appropriate storage capability to store large datasets and have fast read/write speeds. This way, the training and assessment of the SWCMS model would be highly efficient with escalations of inaccuracies through increased computational skills to address water quality problems. In areas with unreliable network connectivity, the SWCMS employs robust data transmission protocols designed to ensure reliable sensor data delivery. Techniques such as data buffering and local storage are used to temporarily hold sensor data when connectivity is lost, ensuring no data are lost. Once the connection is restored, the buffered data are transmitted in batches. Additionally, low-power wide-area networks and protocols like message queuing telemetry transport are utilized for efficient and reliable data transmission, even in low-bandwidth environments, ensuring continuous monitoring and timely updates to the system.

Table 3 exhibits the hyperparameters that have been chosen according to the most used values in similar applications and fine-tuned in compliance with the specifications of the SWCMS. All these make sure that the models operate optimally and increase their effectiveness.

**Table 3** | Empirical configuration to evaluate SWCMS

Approaches	Hyperparameter	Optimal range/values
CNN	Number of conv layers	3
	Filters per layer	64
	Kernel size	(3, 3)
	Pooling size	(2, 2)
	Activation function	Rectified linear init (ReLU)
	Dropout	0.5
	Batch size	32
	Optimizer	Adam; <i>Adjovu et al. (2023)</i>
LSTM	Learning rate	0.001
	Number of LSTM layers	2
	Units per LSTM layer	50
	Activation function	tanh
	Dropout rate	0.3
	Batch size	32
	Optimizer	Adam
RF	Learning rate	0.001
	Number of trees	100
	Maximum depth	10
	Minimum samples split	2
	Minimum samples leaf	1
GBM	Bootstrap	True
	Learning rate	0.1
	Number of estimators	200
	Maximum depth	3
	Minimum samples split	2
SDAE	Minimum samples leaf	1
	Subsample	0.8
	Number of layers	3
	Units per layer	64
	Noise level	0.2
	Activation function	ReLU
	Dropout rate	0.3
Final detection	Batch size	32
	Optimizer	Adam
	Learning rate	0.001
	Weight for GBM ( $\alpha$ )	0.6
Final detection	Weight for SDAE ( $1 - \alpha$ )	0.4
	Epochs	100
	Training:Testing	80:20

**Table 4** | Detection samples of SWCMS for various water quality parameters

Sample ID	pH	Sodium	Magnesium	Calcium	Chloride	Potassium	Carbonate	Sulphate	TDS	TH	EC ( $\mu\text{S/cm}$ )	WQI	Detection flag
		(in mg/L)											
S1	7.49	168.17	10.13	93.63	189.46	0.66	56.72	65.51	1,835.52	188.23	1,802.43	81.19	0
S2	6.87	134.37	1.48	80.23	<b>122.25</b>	3.51	<b>434.67</b>	118.76	771.67	102.92	1,485.71	69.11	<b>1</b>
S3	7.23	<b>158.22</b>	15.85	36.56	112.94	4.86	<b>376.65</b>	83.93	1,740.36	207.04	<b>2,227.81</b>	44.23	<b>1</b>
S4	6.71	13.22	37.09	36.79	43.16	7.49	489.89	36.11	926.08	177.69	2,127.57	71.79	0
S5	<b>8.22</b>	162.68	23.76	49.72	<b>173.91</b>	6.42	268.83	118.41	768.52	226.34	612.42	42.53	<b>1</b>
S6	7.53	120.57	33.48	54.88	50.69	1.04	256.21	64.46	1,039.66	132.66	2,514.94	76.13	0
S7	7.14	54.43	29.43	5.27	117.16	8.88	88.98	13.09	1,782.54	418.19	1,873.41	61.73	0
S8	<b>7.86</b>	73.44	7.09	<b>97.47</b>	25.99	6.52	195.62	88.82	651.44	<b>487.01</b>	1,554.16	49.47	<b>1</b>
S9	7.31	142.77	23.67	94.53	103.02	7.51	384.21	209.42	1,411.83	303.21	667.76	74.84	0
⋮													
S100	<b>8.21</b>	<b>170.94</b>	22.71	78.41	<b>148.59</b>	2.63	480.34	155.57	1,446.37	384.22	1567.85	<b>93.38</b>	<b>1</b>

Table 4 lists a few detection samples of the SWCMS of different water quality indexes as well as the flag of detection as normal (0) or alarm (1). When evaluating the results of the detection concerning the SWCMS, it was confirmed that the system is able to detect problems with water quality. The samples that have a detection flag of 1 demonstrate the concerning parameter values that are out of the ordinary range, which may be a sign of water pollution. For instance, the high chloride and carbonate concentrations in Sample S2, the high sodium and high EC in Sample S3, and the extremely low/high pH in Sample S5 and Sample 100 are interesting. Besides, sample S8 is critical for high levels of sulphate and TDS (Hicks *et al.* 2022). These deviations are, however, necessary for giving an alert and causing awareness as soon as possible to consumers concerning water quality problems. The observed attributes are quite explicit and clearly marked elevated sulphate, TDS, as well as the filters' ability to determine the presence of specific disturbances, only proving that the system is rather reliable and stable for real-time water quality detection.

The proposed CA approach from the SWCMS is compared with the following existing methods for performance evaluations: MCN-LSTM, LSTM, gradient booster regressor (GBR), FL-DT, and RF. Table 5 reflects the comparative performance analysis (Psaros *et al.* 2023) of the existing methods, which suggests the marked superiority of the CA in the SWCMS. The proposed approach for people re-identification attains an average accuracy of 96.0%, precision of 95.9%, and recall of 95.8% with an F1 score of 95.8%. Thus, CA (SWCMS) presents the best performance, meaning that it is the most suitable for processing water quality data with high precision and reliability. MCN-LSTM is also accurate, achieving an accuracy of 92.5% and a precision of 91.8%, although it is computationally expensive. Thus, the presented LSTM and RF models can be considered satisfactory when it comes to a balance between performance and resource usage; yet again, they are not as efficient as CA (SWCMS). While subtler and not as computational, the performance metrics for the two models are lower; notably, the FL-DT is the lowest. Thus, based on a balance between the computational costs and high-performance characteristics, it is possible to conclude that CA (SWCMS) will be the most appropriate for water quality analysis.

The use of several lower-level and higher-level models in CA is naturally accompanied by higher requirements for computational complexity (Rogers & Louis 2005) and increased time consumption. Furthermore, combining outputs of several models and conditions to guarantee real-time performance increases the workload; therefore, the inference time is impacted. Such computation limitation can be addressed by introducing refined parallel computing in a distributed environment.

Table 6 shows that the MCN-LSTM has great accuracy with a significant training time of about 6.5 h among the models and a moderate amount of memory, about 3.2 GB, and the inference time was fairly efficient, estimated to be 0.05 s/sample. LSTM is favourable in training time with a duration of 5 h, a memory size of 2.8 GB, and a slightly faster inference time of 0.03 s/sample. It also has slightly lower accuracy compared to the MCN-LSTM model. Thus, the GBR model, with its training time of 3.5 hours, consumed less memory (1.5 GB) and had an efficient inference time of 0.02 s/sample, but the algorithms revealed lower performances than others. Overall, the FL-DT model is the least complex, with a training time of 2 h and an inference time of 0.01 s/sample, and the additional memory requirement is approximately 1.2 GB of memory, yet it provided the lowest performance statistics when the other approaches provided better results. The RF model consumes 3 h of time to train with 1.7 GB of memory, while the inference is very efficient, 0.02 s/sample, but still lower than CA computational performance-wise. Despite committing the highest training time (8 h) and memory usage (3.5 GB), the CA (SWCMS) supersedes all models in all measures of accuracy, precision, recall, and F1 score, with an inference time of 0.06 s/sample, which makes them the best option for monitoring the quality of water in the most efficient way.

**Table 5** | Comparative performance evaluation and analysis of CA with existing approaches

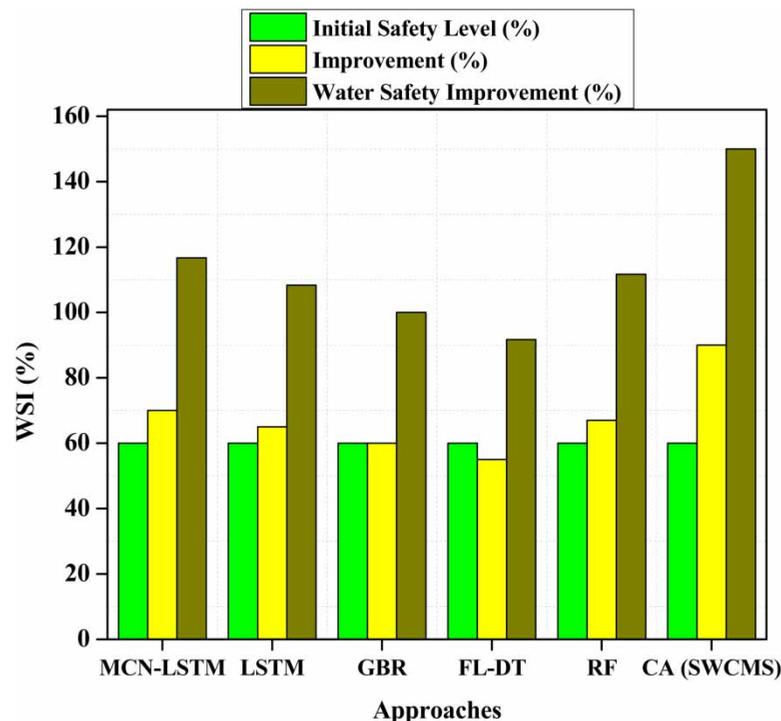
Method	Performance metrics (%)			
	Accuracy	Precision	Recall	F1 score
MCN – LSTM	92.5	91.8	91.5	91.6
LSTM	90.2	89.7	89.3	89.5
GBR	88.7	88	87.5	87.7
FL-DT	85.4	84.8	84.2	84.5
RF	89.9	89.3	89	89.1
CA (SWCMS)	96	95.9	95.8	95.8

**Table 6** | Comparison of computation efficiency

Method	Training time (in h)	Inference time (in s/sample)	Memory usage (in GB)
MCN-LSTM	6.5	0.05	3.2
LSTM	5	0.03	2.8
GBR	3.5	0.02	1.5
FL-DT	2	0.01	1.2
RF	3	0.02	1.7
CA (SWCMS)	8	0.06	3.5

Water safety improvement (WSI) (Ahmed *et al.* 2022) quantifies the beneficence of the system in improving the safety of the water that is used for drinking. This metric assesses the identification and management of possible threats to water quality; attainment of lower risks associated with polluted water is hereby postulated in the system. The percentage increase in water safety the system brings is determined by comparing the detection and prevention of water quality due to the system. This improvement is usually assessed with specific reference to the way the system identifies and reacts to incidences of poor water quality in the period before it was implemented and compares it to the period after it was implemented. The additional percentage increase in water safety is derived from comparing the enhancement of the system to the first degree of protection achieved.

A comparison of WSI accomplished by various methods is presented in Figure 2, with the initial safety level being 60%. The proposed CA in SWCMS shows the highest improvement, improving water safety by 90%, and overall by 150%. This enormous improvement supports the system's ability to identify and contain pollution threats, thus projecting it as the most efficient technique compared to the others that were compared in this research. The MCN-LSTM-based method also records a very significant improvement of a remarkable 70% and gives an overall safety improvement of 116.67%, which shows its effectiveness in terms of safety in processing extensive and intricate data on water quality. LSTM comes next with a 65%

**Figure 2** | Analysis of WSI.

improvement and a 108.33% overall safety enhancement. Such effectiveness proved the technology's ability to perform temporal data analysis. GBR reaches a 60% improvement, which is equal to 100% on the whole, which highlights the result of high efficiency in the tasks of regression. In the case of getting the marked improvement from low accuracy, FL-DT, with the lowest improvement of 55%, yields the highest accuracy of 91.67% overall increase demonstrated, which is not very high and confirms the fact that the tool does not significantly affect water safety. RF provides a balanced performance with an enhancement of 67%, and an overall improvement of 111.67% improvement was recorded, showing the efficacy of the model. To sum up, all the methods are useful in enhancing water safety, but CA (SWCMS) is the most efficient one, pointing at the highest increase in the index and providing the most substantial confidence in the quality management of water.

Consumer awareness enhancement (CAE) (Javanbakht-Sheikhahmad *et al.* 2024) assesses the extent to which the system creates awareness among consumers about the quality of the drinking water they consume. This measure evaluates the extent to which the system updates users on the present water situation and the danger of water-borne diseases thus allowing the users to act accordingly.

The relation of CAE of each method is shown in Figure 3, where the initial awareness level is set to be 50%. Specifically for SWCMS, the proposed CA shows the highest improvement by raising consumer awareness by 85%, which leads to 92.5% overall enhancement. Such a significant rise epitomizes the efficiency of the system in averting water-borne diseases through the supply of timely and accurate water quality information, thus enhancing consumer awareness. Hence, the MCN-LSTM has achieved a 60% improvement, resulting in an 80% boost on the whole, proving the model's efficiency in processing large and intricate data and providing the necessary information to the users.

The percentage of increased improvement in the LSTM method is 55% higher and, therefore, makes a 77.5% overall improvement, proving its prowess in dealing with sequential data and increasing the user's level of knowledge. GBR gets an improvement of 50%, which gives an overall improvement of 75%, showing the effectiveness of GBR in regression problems but with comparatively less change in consumers' awareness. Thus, the improvement in the condition of FL-DT is 45% lower compared to other approaches and attains an overall awareness of 72.5%. This conclusion underlines that the effect of Facebook advertising on the increase of consumers' awareness is rather low. For RF, an impactful level of convergence was displayed, with 58% improvement and 79% overall enhancement. In other words, all the methods help in increasing consumer awareness, but CA (SWCMS) is the most effective one, providing a higher degree of enhancement

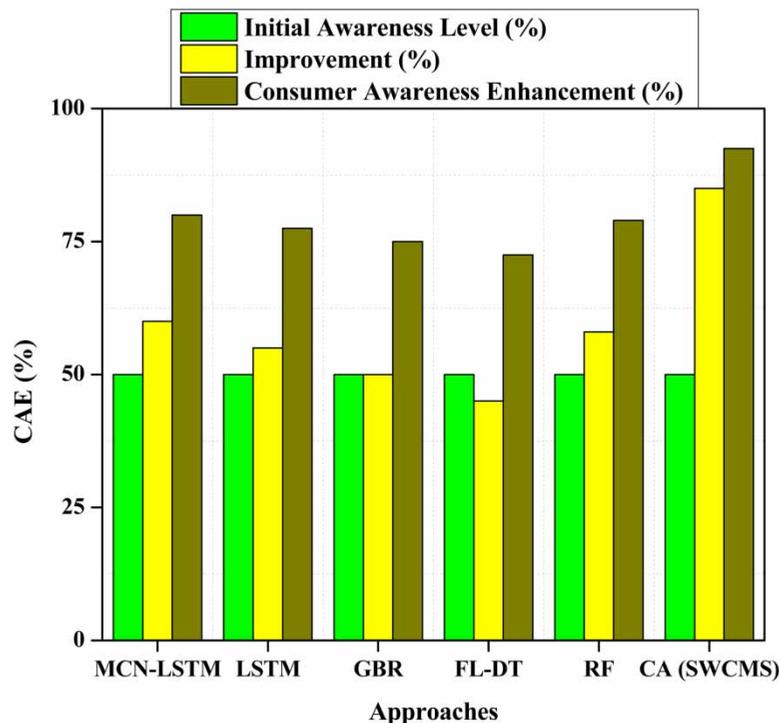


Figure 3 | Analysis of CAE.

and guaranteeing a higher level of awareness and timely response to water quality concerns. With real-time monitoring of the water, the SWCMS helps to prevent water-borne diseases and increase the safety of water by detecting problems as they occur. In the cost-effectiveness model view, SWCMS may be costly initially and in operations because of the technology and computation burdens necessary to support educational models at their most complex and accurate; however, these costs are far outshaded by gains. Due to its availability to bring timely alerts or shoot down the frequency of water-related health hazards, the system has the edge over conventional water monitoring practices and is instrumental in containing healthcare costs and improving public health outputs. With this, SWCMS offers more benefits and is cheaper than the current working solution in the long run.

Thus, the rise in consumption awareness translates to:

- Increased awareness among consumers on water consumption matches the firm's objectives.
- Spontaneous reactions to cases of water quality concerns have been reported.
- Strengthened health of the public due to increased knowledge and preventive actions.

The technical consequences of the CA at the level of the SWCMS architecture are dramatic because it assumes a complex decisional model that is able to strengthen the reliability of water quality assessments. From this point of view, consistently applying the lower-level models like CNNs for extracting spatial features and RNNs with LSTM for capturing temporal dependencies and incorporating the RF for the feature selection, the system forms an exhaustive feature set. This feature set passes through the higher-level models, such as the GBM and the SDAE, to improve the prediction and boost the interactivity within the dataset. Each of these higher-level models' outputs accumulated creates realistic and robust forecasts. These predictions are then reported and brought to consumers' awareness in real time via a mobile application, thereby allowing for timely prompt action. The CA approach guarantees that every layer utilizes the advantages of each model, thus developing a strong and effective structure for assessing and perceiving the condition of drinking water, which will contribute to the betterment of public health and safety. There are many predictions about the effectiveness of the proposed SWCMS, especially in terms of scalability, as well as inter-linkage between urban and rural domains. In urban settings, the data processing of big volumes of real-time data makes it easier to monitor water quality in wide networks, hence creating public health and safety. Despite the potential coverage gap in rural areas with regard to water quality information due to usage of the SWCMS, concerned people can gain important and timely information to enhance water safety and awareness, with increased comparability between rural and urban water management.

Therefore, the facilitation that comes with the SWCMS has far-reaching consequences on the health of the population and security within the societies in question. Based on sophisticated sensor solutions and data analysis techniques, SWCMS enhances the real-time, automatic evaluation of water quality, thus greatly enhancing water quality and consumers' health. Due to the timely nature of the alerts and detailed information, the consumers are in a position to make appropriate decisions on water quality, hence avoiding incidences of water-borne illnesses. Furthermore, based on the findings of the proposed study, the high performance of the CA within SWCMS proves the existence of higher accuracy, precision, and effectiveness in implementing water quality management systems. In essence, consumers benefit from safer and improved quality of drinking water through the use of SWCMS, which also promotes the health of the public through awareness of safe drinking water and streaks off diseases associated with impure water.

However, the CA approach in SWCMS has some limitations, even though it has its advantages. Due to the highly complex structure, as well as the incorporation of multiple models at different levels, there are likely to be higher computational requirements, as well as higher time in training when compared with more simple models implemented at individual levels, especially at socio-hydrology management (Javanbakht-Sheikhahmad *et al.* 2024). Also, the requirement of broad datasets to train the rich models efficiently may be problematic in data collection and computation. Table 7 depicts the significant comparative discussion with existing systems based on the key comparative factors.

The training phase of the SWCMS, which uses CG, has a certain computational complexity and high memory consumption because these models are multilayered. CNNs and LSTMs consume a large amount of computational resources while dealing with spatial and temporal data in Layer 1, and the RF model also takes in terms of memory as it is a kind of ensemble of decision trees. The intermediate fusion layer integrates outputs from these models, thus raising dimensionality and memory requirements. In the higher level models such as GBM and SDAE, further improvements in the prediction are made at the cost of time taken for the computation. Feature engineering methods, including spatial pattern analysis by the CNN, temporal analysis by the LSTMs, and feature selection by the RF, are more important in improving the model

**Table 7** | Comparison between the proposed SWCMS model and other WQM techniques

Comparative factors	Limitations	Performance	Connectivity	Sensor system	Data processing	Parameters monitored	Core objective
SWCMS	Computation cost	Detection accuracy >95%	Mobile application for real-time alerts	High-performance real-time sensors	Mean imputation and Z-score normalization	pH, turbidity, temperature, chemicals	Real-time WQM
Lakshmikantha <i>et al.</i> (2021)	No clear data collection protocol	NA	NA	Low-cost sensors	Standardized data analysis	NA	SHM of pipes
Garrido-Momparler & Peris (2022)	Dependence on internet connection	Low cost, effective	Internet connection required	Sensors with Arduino microcontroller	Cloud server evaluation	pH, turbidity, conductivity, temp	IoT-based WQM
El-Shafeiy <i>et al.</i> (2023)	Power consumption, price issues	High frequency and accuracy	IoT infra and cloud computing	Smart sensors	Cloud computing	NA	Smart sensor connectivity
Syrmos <i>et al.</i> (2023)	Computationally intensive	Early, accurate signals	Network of IoT-based sensors	IoT-based sensors	Deep learning (MCN-LSTM)	NA	Outlier detection in water data
Chinnappan <i>et al.</i> (2023)	Sensor calibration and maintenance	Real-time monitoring	LoRaWAN	Flow meters, water quality sensors	ML predictive models	Consumption rates, quality features	IoT architecture for real-time WQM
Shahra & Wu (2023)	Data quality and availability	Constant supply of safe water	Cloud connectivity	Sensors with Raspberry Pi	FL-DT	Chlorine level, temperature, flow	Real-time chlorine level checks
Rana <i>et al.</i> (2023)	High computational resources	Shortest detection time, wide coverage	NA	NA	EA	NA	Optimal sensor placement
Jallow (2024), Shams <i>et al.</i> (2024)	High computational complexity	High success rates (MAE, MSE)	NA	NA	ANN, LSTM	Dissolved oxygen, temp, conductivity, pH, turbidity, TDS, chlorides	Water quality prediction and monitoring
Ruiz-Moreno <i>et al.</i> (2023)	Conceptual stage	Conceptual stage	Cloud-based	Microscopic cameras, bioluminescent sensors	TensorFlow Javascript	ATP, nitrate	Assessing water quality with ML

performance as they determine which of the features to be used at the final prediction stage have been preprocessed. The highly complicated relationships between the features were learned in the GBM to enhance the prediction accuracy, while the SDAE focused on creating better representations of the features by removing noise, thereby leaving only the most informative patterns for the final prediction.

## 5. CONCLUSION AND FUTURE RESEARCH

Some of the metrics used in assessing the SWCMS performance to the 95% detection accuracy include precision, recall, F1 score, training time, inference time per sample, memory usage, WSI, and CAE. Precision measures how correct positive predictions are in regard to the actual positive classification. Recall measures the precision of the model while identifying the total number of positive cases, and the F1 score combines the measures of both precision and recall into a single value. Training time is the time of training the model efficiently, while inference time shows how faster the model is in terms of real-time prediction. Memory usage shows how the system utilizes the system resources, while WSI and CAE quantify the enhancement in water safety and consumer awareness that the system brings. Combined, these metrics build a solid and relatively efficient system that envisages a high degree of detection accuracy and high overall effectiveness of the SWCMS.

In conclusion, the SWCMS, with its CA, improves WQM through the latest AI models, such as CNNs, LSTMs, RF, GBMs, and SDAEs. Another benefit of the system is the real-time identification of problems in water quality, which consequently raises water safety by 90% and consumer sensitivity by 92.5%. Even though training the CA model is costlier than training the usual models, time-consuming, and requires more memory, CA can overcome these drawbacks with parallel processing, fast hardware, optimized models, and data processing. The complex strategy of CA proves that the combination of single, distinctive models and time and space data provides more accurate and successful water quality evaluation. Through the provision of timely alerts or detailed information to the consumers, the users of SWCMS are in a position to protect themselves, and thus, the welfare of society is boosted. Gradual implementation helps identify technical or operational issues that may arise, hence avoiding massive failures. Additionally, updating and maintaining the ML models from time to time keeps SWCMS equipped for water quality threats and brings optimum results in the long run. Due to this development, the system's effectiveness from a technical perspective, as well as its utility to stakeholders, makes it one of the models for dealing with water quality.

The future developments of SWCMS are as follows: The SWCMS is planned to be enhanced by integrating concurrent computing frameworks to manage large datasets more effectively and by adapting higher-grade computing hardware to accelerate processing. Additionally, future developments will explore the use of methodologies like transfer learning and quantization to further improve model accuracy while optimizing resource utilization. These advancements aim to ensure the system remains scalable and efficient as it evolves. Further, improving model architectures using some of the methodologies, such as transfer learning and quantization, could improve the accuracy while at the same time optimizing the utilization of resources.

## ACKNOWLEDGEMENTS

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education, Saudi Arabia, for funding this research (IFKSUOR3-176-7).

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## CONFLICT OF INTEREST

The authors declare there is no conflict.

## REFERENCES

- Adjovu, G. E., Stephen, H., James, D. & Ahmad, S. (2023) Measurement of total dissolved solids and total suspended solids in water systems: A review of the issues, conventional, and remote sensing techniques, *Remote Sensing*, **15** (14), 3534.
- Ahmed, T., Sipra, H., Zahir, M., Ahmad, A. & Ahmed, M. (2022) Consumer perception and behavior toward water supply, demand, water tariff, water quality, and willingness-to-pay: A cross sectional study, *Water Resources Management*, **36** (4), 1339–1354.
- Ajayi, O., Bagula, A. & Maluleke, H. (2022a) Dataset for assessing water quality for drinking and irrigation purposes using machine learning models. Available at: <https://dx.doi.org/10.21227/trcf-1s03>.

- Ajayi, O. O., Bagula, A. B., Maluleke, H. C., Gaffoor, Z., Jovanovic, N. & Pietersen, K. C. (2022b) [WaterNet: A network for monitoring and assessing water quality for drinking and irrigation purposes](#), *IEEE Access*, **10**, 48318–48337.
- Borup, D., Christensen, B. J., Mühlbach, N. S. & Nielsen, M. S. (2023) [Targeting predictors in random forest regression](#), *International Journal of Forecasting*, **39** (2), 841–868.
- Chinnappan, C. V., John William, A. D., Nidamanuri, S. K. C., Jayalakshmi, S., Bogani, R., Thanapal, P., Syed, S., Venkateswarlu, B. & Syed Masood, J. A. I. (2023) [IoT-enabled chlorine level assessment and prediction in water monitoring system using machine learning](#), *Electronics*, **12** (6), 1458.
- Dong, Z., Hou, K., Meng, H., Yu, X. & Jia, H. (2022) [Data-driven power system reliability evaluation based on stacked denoising auto-encoders](#), *Energy Reports*, **8**, 920–927.
- El-Shafeiy, E., Alsabaan, M., Ibrahim, M. I. & Elwahsh, H. (2023) [Real-time anomaly detection for water quality sensor monitoring based on multivariate deep learning technique](#), *Sensors*, **23** (20), 8613.
- Garrido-Momparler, V. & Peris, M. (2022) [Smart sensors in environmental/water quality monitoring using IoT and cloud services](#), *Trends in Environmental Analytical Chemistry*, **35**, e00173.
- Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P. & Parasa, S. (2022) [On evaluation metrics for medical applications of artificial intelligence](#), *Scientific Reports*, **12** (1), 5979.
- Huang, T., Zhang, Q., Tang, X., Zhao, S. & Lu, X. (2022) [A novel fault diagnosis method based on CNN and LSTM and its application in fault diagnosis for complex systems](#), *Artificial Intelligence Review*, **55** (2), 1289–1315.
- Jallow, C. B. (2024) [LOWEL AI using machine learning algorithms and sensors to detect harmful bacteria present in water](#), *Journal of Agricultural, Earth and Environmental Sciences*, **3** (3), 1–6.
- Javanbakht-Sheikhahmad, F., Rostami, F., Azadi, H., Veisi, H., Amiri, F. & Witlox, F. (2024) [Agricultural water resource management in the socio-hydrology: A framework for using system dynamics simulation](#), *Water Resources Management*, **38** (8), 2753–2772.
- Lakshmikantha, V., Hiriyannagowda, A., Manjunath, A., Patted, A., Basavaiah, J. & Anthony, A. A. (2021) [Iot based smart water quality monitoring system](#), *Global Transitions Proceedings*, **2** (2), 181–186.
- Lowe, M., Qin, R. & Mao, X. (2022) [A review on machine learning, artificial intelligence, and smart technology in water treatment and monitoring](#), *Water*, **14** (9), 1384.
- Manjakkal, L., Mitra, S., Petillot, Y. R., Shutler, J., Scott, E. M., Willander, M. & Dahiya, R. (2021) [Connected sensors, innovative sensor deployment, and intelligent data analysis for online water quality monitoring](#), *IEEE Internet of Things Journal*, **8** (18), 13805–13824.
- Manoharan, A., Begam, K. M., Aparow, V. R. & Sooriamorthy, D. (2022) [Artificial neural networks, gradient boosting and support vector machines for electric vehicle battery state estimation: A review](#), *Journal of Energy Storage*, **55**, 105384.
- Mishra, R. K. (2023) [Fresh water availability and its global challenge](#), *British Journal of Multidisciplinary and Advanced Studies*, **4** (3), 1–78.
- Nasir, N., Kansal, A., Alshaltone, O., Barneih, F., Sameer, M., Shanableh, A. & Al-Shamma'a, A. (2022) [Water quality classification using machine learning algorithms](#), *Journal of Water Process Engineering*, **48**, 102920.
- Pсарos, A. F., Meng, X., Zou, Z., Guo, L. & Karniadakis, G. E. (2023) [Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons](#), *Journal of Computational Physics*, **477**, 111902.
- Rana, R., Kalia, A., Boora, A., Alfaisal, F. M., Alharbi, R. S., Berwal, P., Alam, S., Khan, M. A. & Qamar, O. (2023) [Artificial intelligence for surface water quality evaluation, monitoring and assessment](#), *Water*, **15** (22), 3919.
- Rogers, J. W. & Louis, G. E. (2005) [A standard efficiency metric for evaluating the performance of community water systems](#), *Journal-American Water Works Association*, **97** (10), 76–86.
- Ruiz-Moreno, S., Gallego, A. J., Sanchez, A. J. & Camacho, E. F. (2023) [A cascade neural network methodology for fault detection and diagnosis in solar thermal plants](#), *Renewable Energy*, **211**, 76–86.
- Seu, K., Kang, M. S. & Lee, H. (2022) [An intelligent missing data imputation techniques: A review](#), *JOIV: International Journal on Informatics Visualization*, **6** (1–2), 278–283.
- Shahra, E. Q. & Wu, W. (2023) [Water contaminants detection using sensor placement approach in smart water networks](#), *Journal of Ambient Intelligence and Humanized Computing*, **14** (5), 4971–4986.
- Shams, M. Y., Elshewey, A. M., El-Kenawy, E. S. M., Ibrahim, A., Talaat, F. M. & Tarek, Z. (2024) [Water quality prediction using machine learning models based on grid search method](#), *Multimedia Tools and Applications*, **83** (12), 35307–35334.
- Silva, J. A. (2022) [Implementation and integration of sustainability in the water industry: A systematic literature review](#), *Sustainability*, **14** (23), 15919.
- Singh, D. & Singh, B. (2022) [Feature wise normalization: An effective way of normalizing data](#), *Pattern Recognition*, **122**, 108307.
- Syrmos, E., Sidiropoulos, V., Bechtsis, D., Stergiopoulos, F., Aivazidou, E., Vrakas, D., Vezinias, P. & Vlahavas, I. (2023) [An intelligent modular water monitoring iot system for real-time quantitative and qualitative measurements](#), *Sustainability*, **15** (3), 2127.
- Talukdar, S., Ahmed, S., Naikoo, M. W., Rahman, A., Mallik, S., Ningthoujam, S., Bera, S. & Ramana, G. V. (2023) [Predicting lake water quality index with sensitivity-uncertainty analysis using deep learning algorithms](#), *Journal of Cleaner Production*, **406**, 136885.
- Wang, Y., Xiao, Z. & Cao, G. (2022) [A convolutional neural network method based on adam optimizer with power-exponential learning rate for bearing fault diagnosis](#), *Journal of Vibroengineering*, **24** (4), 666–678.
- Xu, H., Berres, A., Liu, Y., Allen-Dumas, M. R. & Sanyal, J. (2022) [An overview of visualization and visual analytics applications in water resources management](#), *Environmental Modelling & Software*, **153**, 105396.