https://doi.org/10.48047/AFJBS.6.14.2024.8124-8134

African Journal of Biological Sciences

Journal homepage: http://www.afjbs.com

Research Paper Open Access

Water Net: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes

Sarakana Sai Leela Nagamani[1], Murty Sivala Vishnu M.Tech, Ph.D.[2]

[1]PG Student, Department of CSE, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh.

[2]Associate Professor, Department of CSE, Aditya Institute of Technology and Management, Tekkali, Andhra Pradesh.

Volume 6, Issue 14, Aug 2024

Received: 15 June 2024

Accepted: 25 July 2024

Published: 15 Aug 2024

doi: 10.48047/AFJBS.6.14.2024.8124-8134

Abstract: Water is essential for the survival of humans, animals, and plants, yet its quality often falls short of suitability for various purposes due to industrialization, mining, pollution, and natural factors altering its composition. Regulatory bodies like the World Health Organization set guidelines for acceptable water quality levels, emphasizing parameters crucial for human consumption and irrigation. Assessing water quality involves laborious sampling, parameter measurement, and adherence to stringent guidelines, presenting challenges. This study proposes a network architecture utilizing LoRa technology to collect real-time water parameter data, aiming to automate the assessment of water suitability for drinking and irrigation using machine learning (ML) tools. Simulations employing Radio Mobile suggested a partial mesh network topology as most effective for the monitoring network, considering land topology. Given the scarcity of large, open datasets for drinking and irrigation water, the study developed usable datasets for ML model training. Three ML models—Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM)—were evaluated for water classification, with LR showing superiority for drinking water and SVM for irrigation. Recursive feature elimination was employed with the ML models to identify the most influential water parameters on classification accuracies.

Keywords: Cyber physical system, LoRa, drinking water, irrigation water, machine learning, water quality index, water monitoring network.

I. INTRODUCTION

Access to water is deemed crucial for human survival, recognized as a fundamental human right and integral to achieving the Sustainable Development Goals (SDGs) set forth by the United Nations in 2015, particularly SDG 6, which emphasizes the universal access to clean water and sanitation. Clean water access not only aligns with SDG 3, promoting good health and well-being, as contaminated water serves as a vector for diseases like cholera and typhoid, leading causes of mortality, particularly among children, in developing regions of Africa and Asia. Additionally, water plays a vital role in agriculture and food production, with approximately

10% of the global population suffering from malnutrition, predominantly in developing nations, where hunger contributes to 45% of infant mortality. Recognizing food security as a critical imperative, SDG 2 aims to end hunger by promoting sustainable agriculture and improving food distribution, highlighting the reliance of food production on water for irrigation and animal consumption. Hence, ensuring the availability and sustainable management of water suitable for agricultural use remains paramount for global food security efforts. Various water sources, such as rivers, streams, rainfall, and groundwater extracted from wells and boreholes, serve both drinking and irrigation purposes. The characteristics of these water sources significantly influence their composition, with natural elements and human-derived chemical pollutants from activities like mining and industrial processes altering water quality. As these contaminated waters are utilized in households and farms for various purposes, including drinking, livestock consumption, and crop irrigation, there's a heightened risk of adverse health effects or even fatalities. Therefore, it is imperative to implement a comprehensive monitoring system to track water quality from its source to its final point of use. Regular water sampling and quality assessments are essential at each monitoring location to ensure suitability for human and animal consumption, irrigation, and domestic or industrial applications. Various models have been developed to evaluate water quality, taking into account a range of parameters including chemical components (such as pH, calcium, oxygen, sulphate levels), microbial content (such as E. coli, rotaviruses, Entamoeba), and physical attributes (temperature and clarity). These models yield a standardized metric known as the Water Quality Index (WQI). Different regions worldwide follow distinct guidelines for calculating WQI. For example, in Europe, standards like the British Columbia Water Quality Index (BCWQI) and the Scottish Research Development Department (SRDD) are utilized, while North America predominantly uses models such as the Canadian Council of Ministers of the Environment Water Quality Index (CCMEWQI) and National Sanitation Foundation Water Quality Index (NSFWQI). In Asia, particularly in India, the Bureau of Indian Standards (BIS) is prominent, while in Africa, notable standards include the South African National Standard for drinking water (SANS 241-1) and the Kenya Bureau of Standards (KEBS). Many of these models are adapted from guidelines established by the World Health Organization (WHO). This study adheres to standards outlined by South Africa and the WHO. Undoubtedly, assessing water parameters across various samples can be a time-consuming and challenging endeavor, requiring strict adherence to protocols for sample collection, transportation, laboratory analysis, and quality assurance. Detailed procedures for these tasks are outlined in references. The outcomes of these processes determine whether a water sample meets potable standards. In our study, we introduce cyber-physical network architecture for real-time monitoring of water parameters throughout a city, complemented by a machine learning model for assessing water portability. Similar to previous works, our focus is primarily on physical and chemical water parameters, omitting biological aspects. This choice stems from our model's sensor-based nature within the Internet of Things (IoT) framework, as physical sensors capable of measuring biological parameters like E. coli presence are currently unavailable. However, we recognize the significance of microbial water parameters and acknowledge that our model could accommodate them by integrating suitable physical sensors, if accessible, or virtual/soft sensors as proposed.

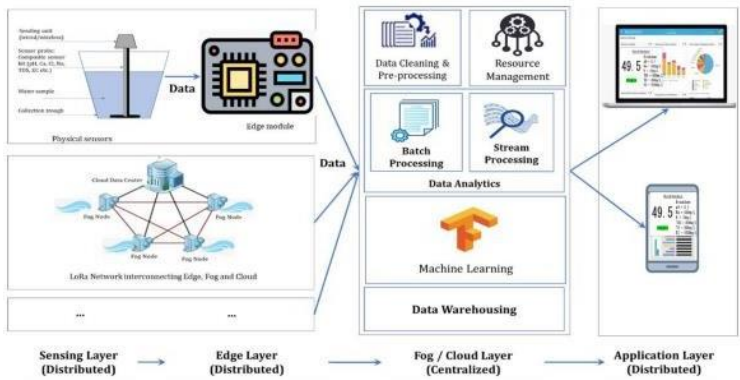


Fig. 1: Conceptual framework for Water Quality Monitoring

Our proposed architecture comprises four layers: the Sensing Layer, Edge Layer, Fog/Cloud Layer, and Application Layer. The Sensing Layer directly interacts with water samples in various sources like rivers, streams, or dams, employing sensor probes equipped with sensors for measuring parameters such as pH, conductivity, turbidity, and temperature. These sensors transmit telemetry data to Fog Nodes (FNs) for processing. The Edge Layer employs low-end processing devices like Raspberry Pi or Arduino to preprocess data and serve as network gateways for transmitting data to FNs via cellular or long-range network solutions. FNs, part of the Fog/Cloud Layer, are distributed cloud computing nodes responsible for classifying water samples using machine learning models. They consider only the most influential parameters due to limited computing power compared to the Cloud, forwarding data to the Cloud Data Centre for advanced analytics when needed. The Cloud Data Centre serves as a high-performance computing infrastructure for data warehousing, analytics, and hosting services. The Application Layer acts as an interface between users and cloud-based software/services, facilitating access to water monitoring tools through mobile and web platforms. Our proposed water monitoring network targets deployment in Cape Town, South Africa, focusing on monitoring water quality in storage dams and treatment plants. Key contributions include establishing a real-time monitoring network tailored to Cape Town's geographical features, curating datasets for training machine learning models, and developing models to identify critical parameters influencing water quality analysis for drinking or irrigation purposes.

II. LITERATURE SURVEY

WIRELESS COMMUNICATION NETWORKS FOR WATER MONITORING

Three studies are referenced: [1] developed a network to monitor water parameters in a Brazilian city, [2] focused on water quality in Mozambique's Limpopo River Basin with 23 monitoring stations, and [3] proposed a model combining genetic algorithms with water quality simulation.

These studies underscore the importance of periodic water sampling and data transfer for decision- making. Lightweight communication protocols like LPWAN technologies (e.g., SigFox, LoRa) are crucial for transmitting data over long distances. Comparison in [4] revealed Ingenu with the longest range, followed by SigFox and LoRa. Studies like [5], [6], and [7] have shown simulation results aligning with real-world tests, supporting the reliability of simulation-based assessments.

ASSESSING WATER POTABILITY

The assessment of drinking water quality commonly relies on the Water Quality Index (WQI), a unitless metric indicating water suitability for human consumption or general usage. Various WQI models exist globally, including the Horton Index, National Sanitation Foundation WQI, CCME WQI, SRDD index, BWQI, Fuzzy Interface system, and MWQI. Uddin et al. reviewed 35 WQI models, highlighting their structural composition, parameters considered, and limitations, with most models deeming a WQI value of at least 50 acceptable. Another study focused on parameter importance, using analytical hierarchical process (AHP) and MACBETH to assign weights to water parameters. Assessing mining impacts on water quality in Bangladesh, one study considered 12 parameters against WHO standards to determine WQI, while another examined urban water resource management across 12 monitoring points, using CCME and Cetesb WQI models. However, a limitation of WQI is its site specificity and use case constraints. Addressing this, a universal WQI model was proposed for South Africa, aggregating 13 parameters with a custom function to classify water samples effectively, irrespective of their source.

ASSESSING WATER QUALITY FOR IRRIGATION

Assessing the quality of irrigation water is crucial for crop farming, as water quality directly impacts crop yield. While classical techniques and indices like the Irrigation Water Quality Index (IWQI) exist, they often focus solely on drinking water or require numerous parameters, making them economically unfeasible for local farmers. Researchers have proposed alternate techniques based on machine learning (ML), such as predicting water quality parameters like Exchangeable Sodium Percentage (ESP), Magnesium Adsorption Ratio (MAR), and Total Dissolved Solids (TDS) using ML models like Adaboost and Random Forest. Other studies have developed models specifically for irrigation water quality assessment, reducing parameters to sodium, chloride, and electrical conductivity (EC), with Random Forest consistently performing well in classification tasks. The IWQI, a widely used index for assessing water quality for irrigation, considers the relative contribution and weight of each water parameter, often following methodologies proposed by organizations like the WHO. However, these indices have limitations, as they mask the specific effects of individual water parameters on soil and plant health. Many studies in this field lack detailed descriptions of network architectures and communication technologies used for water monitoring, and the application of ML models for water quality interpretation, particularly in developing nations, remains relatively unexplored.

This study aims to address these research gaps by proposing a comprehensive approach to water monitoring and exploring the potential of ML in interpreting water sample analyses economically.

III. METHODOLOGY

Our implementation process was divided into two phases: Phase A, focusing on WaterNet, the water monitoring network, and Phase B, assessing water quality based on data from WaterNet. In Phase A, we simulated WaterNet using Google Maps, Topographic-map.com, and Radio Mobile software. Google Maps provides real-time location services, while Topographic-map.com offers detailed geographical landscape information. Radio Mobile, a network planning tool for simulating radio frequency propagation, is particularly suited for Cape Town's uneven geography, which presents challenges for direct line-of-sight radio propagation. We created a custom map in Google Maps with relevant points of interest, imported it into Radio Mobile as a KML file, and built a two-layer hierarchical network model. The lower level connected dams to their respective water treatment plants (WTPs) using a LoRa network with specific configurations for frequency, transmission power, receiver threshold, and antenna height. The higher level connected the WTPs to the ILLIFU Cloud data center using a 2.4 GHz LoRa network with distinct configurations. Figure 5 illustrates the two-layer Cyber Physical hierarchical network of WaterNet, where X-GWs are the gateways at each dam and FN1 to FN7 are the WTPs hosting the Fog Nodes. To extend the range, we used a high spreading factor of 12, which results in lower data rates but is acceptable for our use case, as we are only transmitting small-sized telemetry data at pre-set intervals.

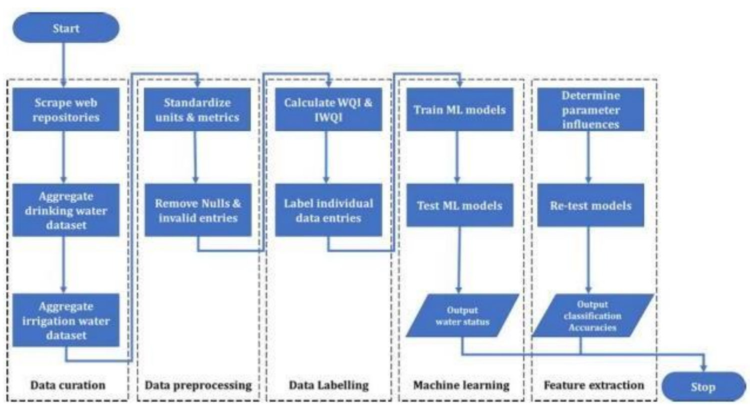


Fig. 2: Process flow for Water Quality Assessment using ML.

With WaterNet established and telemetry data on water parameters being sent to the Fog and Cloud, data analysis and machine learning (ML) can now be employed to assess the water quality at each dam. For this study, we curated data from various sources to simulate parameters received from WaterNet. The simulations presented here focus on using ML to evaluate water quality for drinking or irrigation purposes. Experimental simulations were conducted on Google Colab, using a Python 3 Google Compute module configured with 12 GB of RAM and a 2.3

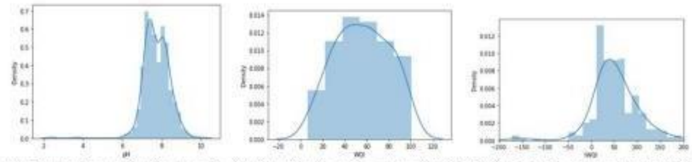
GHz 2-core Intel Xeon CPU. We utilized Sci-Kit Learn for the ML models, Pandas and NumPy for data manipulation, and Matplotlib for data visualization. The dataset was split into 84% training data and 16% test data. Three ML models were considered and compared using five metrics: accuracy, true positive (TP), false positive (FP), false negative (FN), and true negative (TN). Of these metrics, accuracy, FP, and FN were the most critical. Accuracy measures a model's classification performance, i.e., the percentage of correctly classified water samples. False positive (FP) measures the percentage of impure water samples misclassified as potable, which is crucial as misclassifying non-potable water as drinkable can have severe health consequences. False negative (FN) measures the percentage of potable water samples incorrectly classified as unsafe for consumption.

As previously mentioned, the second objective of this work is to use machine learning (ML) models to automatically classify water samples. We selected the 11 most common water parameters from our dataset sources: pH, sodium, magnesium, calcium, chloride, potassium, sulphate, carbonate, TDS, EC, and TH. The three ML classification models considered were Random Forest (RF), Logistic Regression (LR), and Support Vector Classifier (SVC). RF combines multiple decision trees and can be used for both regression and classification. As a classifier, it outputs the "majority vote" from all individual trees and generally avoids overfitting on training data due to bagging and random feature selection. LR models the probability of an event occurring (the dependent variable) based on one or more independent variables. It is well-suited for binary output probabilities (True or False, 1 or 0) and does not require a linear relationship between the variables. In this work, the 11 features were the independent variables, and potability (1 or 0) was the dependent variable. SVC, a form of Support Vector Machine, is a nonlinear solver for classification and regression problems, performing well with smaller datasets. It seeks to draw a hyperplane that maximizes the distance from support vectors of each class while minimizing data separation errors. We used a linear kernel for our SVC model. Water assessment in this study is treated as a classification problem with the primary goal of categorizing water samples as "fit for use" or not. LR was chosen for its suitability for binary classification problems using the sigmoid function, though it is susceptible to outliers. SVC was selected for its robustness to outliers and effectiveness with smaller datasets, which is relevant to our study. RF was considered for its ability to handle datasets of varying sizes and mixed feature sets and its generally faster performance compared to SVC. These three ML models were chosen because they represent different types of ML approaches: LR is based on statistical regression analysis, SVC on data geometry, and RF on ensemble learning.

IV. RESULTS

DETERMINING PARAMETER INFLUENCE IN IRRIGATION WATER

Figure 6 presents the results of recursive feature elimination (RFE) combined with Logistic Regression (RFE+LR), Random Forest (RFE+RF), and Support Vector Classifier (RFE+SVC) on the irrigation water dataset. It indicates that SSP had the least impact on the classification accuracy of the models, while RSC was the most influential feature.



(a) Distribution of pH values in drinking water dataset (b) Distribution of calculated WQI values in drinking water dataset (c) Distribution of calculated IWQI values in irrigation water dataset

Fig. 3: Density distribution curves for select water parameters.

TABLE 1. Parameter influence on classification accuracies of ML models for drinking water.

Excluded Parameters	LR Accuracies (%)	RF Accuracies (%)	SVC Accuracies (%)
pH & TH	99.22	98.45	97.67
Carbonate & Na	96.90	98.45	96.90
Mg & Na	96.12	98.45	96.90
Chloride & Ca	96.12	94.57	95.35
Carbonate & Ca	96.12	94.57	95.35
Carbonate & Chloride	96.12	94.57	94.57
TDS & Na	96.12	98.45	97.67
TDS & Ca	96.12	93.80	94.57
pH & Mg	95.35	96.12	94.57
Chloride & SO4	95.35	95.35	95.35
pH & Ca	94.57	95.35	93.80
pH & Chloride	94.57	95.35	94.57
pH & SO4	94.57	96.12	96.12
Chloride & Na	94.57	94.57	94.57
Chloride & Mg	94.57	96.12	94.57
pH & Carbonate	93.80	96.12	96.12
Mg & K	93.02	93.80	93.02
Carbonate & K	93.02	92.25	93.02
TDS & K	93.02	92.25	93.02
TH & Na	93.02	98.45	93.80
TH & Mg	93.02	96.12	93.02
pH & K	92.25	95.35	94.57
TH & Ca	92.25	93.80	93.02
TH & Chloride	92.25	93.02	93.02
TH & K	92.25	93.02	93.02
Chloride & TDS	91.47	94.57	92.25
Chloride & EC	90.70	93.80	90.70
TH & Carbonate	90.70	93.80	90.70
pH & TDS	89.92	94.57	91.47
pH & EC	89.92	93.80	90.70
Mg & Carbonate	89.15	93.80	89.15
Mg & SO4	89.15	93.80	89.92
TDS & SO4	89.15	89.92	89.92
TH & SO4	89.15	92.25	88.37
Carbonate & TDS	88.37	92.25	89.15
Mg & TDS	87.60	90.70	89.15
Carbonate & EC	85.27	90.70	86.05
TH & TDS	85.27	89.15	85.27
Mg & EC	84.50	84.50	83.72
TH & EC	83.72	86.05	83.72

TABLE 2. Result of model comparison using all features on irrigation water dataset.

	Model	Accuracy (%)	True Positive (%)	False Positive (%)	False Negative (%)	True Negative (%)
1	RF	94.44	91.67	8.33	2.78	97.22
2	LR	91.67	94.44	5.56	11.11	88.89
3	SVC	93.06	94.44	5.50	8.33	91.67

SAR and Na were also notably influential across all models. EC showed significant influence with RFE+LR and RFE+SVC but not with RFE+RF, whereas the influence of Na was the opposite.

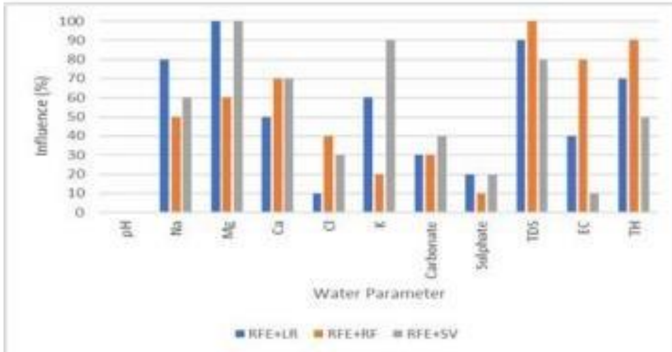


Fig. 4: Parameters influencing the classification accuracy of drinking water

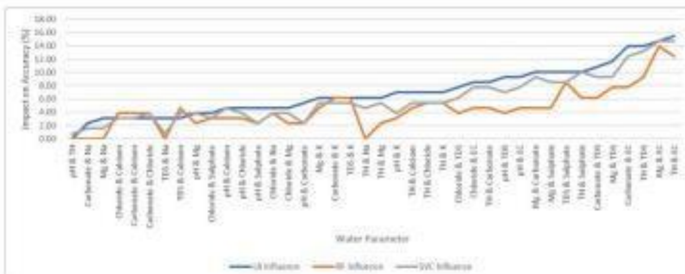


Fig. 5: Impact of various parameters on classification accuracies of drinking water.

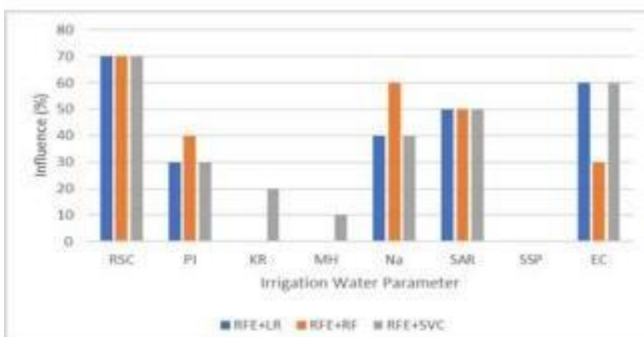


Fig. 6: Parameter influencing the classification accuracy of irrigation water.

These differing influences likely contribute to the lower false positive rates observed with LR and SVC compared to RF, and the lower false negative rates of RF compared to LR and SVC, as shown in Table 2. Table 3 summarizes the results of the top 20 combinations of parameters affecting the performance of LR, RF, and SVC on the irrigation water dataset.

TABLE 3. Parameter influence on classification accuracies of ML models for irrigation water

Excluded Parameters	LR Accuracies (%)	RF Accuracies (%)	SVC Accuracies (%)
SSP & SAR	93.06	93.06	95.83
SSP & PI	90.28	91.67	91.67
SSP & KR	90.28	93.06	93.06
SSP & MH	90.28	93.06	93.06
SSP & Na	88.89	93.06	93.06
KR & RSC	83.33	90.28	79.17
KR & PI	83.33	90.28	79.17
RSC & KR	81.94	90.28	86.11
RSC & MH	81.94	90.28	84.72
Na & RSC	81.94	90.28	86.11
Na & PI	81.94	88.89	80.56
Na & KR	81.94	88.89	80.56
Na & MH	81.94	87.50	81.94
RSC & Na	79.17	91.67	79.17
KR & Na	76.39	87.50	76.39
Na & SSP	76.39	95.83	93.06
RSC & SAR	73.61	87.50	66.67
RSC & SSP	68.06	90.28	84.72
KR & SAR	58.33	77.78	68.06
KR & SSP	58.33	91.67	90.28

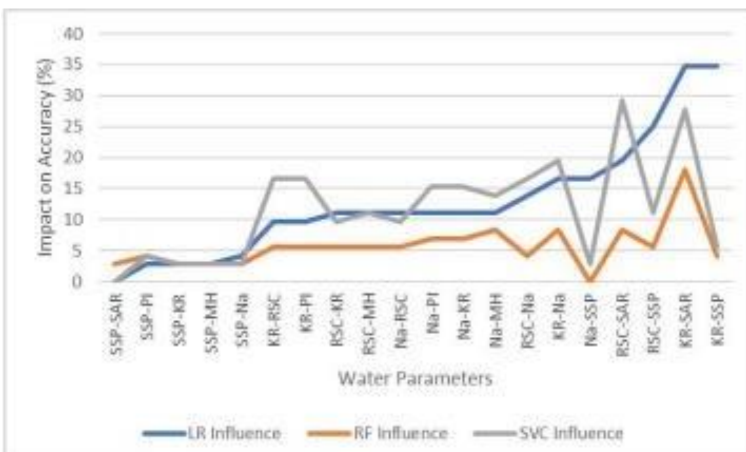


Fig. 7: Impact of various parameters on classification accuracies of irrigation water.

The results in Figure 6 are corroborated by those in Table 3 and Figure 7. This confirms that SSP has the least impact on the classification accuracy of the three ML models considered, while KR and RSC are the most influential parameters for irrigation water.

The WaterNet network architecture demonstrated its effectiveness in real-time monitoring of water parameters using LoRa technology. Logistic Regression (LR) was highly effective for classifying drinking water, whereas Support Vector Machine (SVM) excelled for irrigation water classification. Recursive feature elimination identified key parameters influencing classification accuracy, with RSC and SAR being particularly critical for irrigation water, and a broader range of parameters impacting drinking water quality.

V. CONCLUSION

This work focused on two main concepts: the proposal of a real-time water monitoring network for gathering data on water parameters and the application of machine learning (ML) models to assess water quality. The developed water monitoring network, based on the LoRa protocol for long-range, low-power data transmission, was tested using the City of Cape Town as a case study. Simulations in Radio Mobile indicated that a partial mesh network topology was most suitable for covering the city. Data from this network would ideally be aggregated on a Cloud server, where ML models can assess the water's suitability for drinking or irrigation. Due to a lack of relevant datasets, two suitable datasets were created and used to train and test three ML models: Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM). Results showed that LR performed best for drinking water, with the highest classification accuracy and lowest false positive and negative values, while SVM was better suited for irrigation water. Additionally, recursive feature elimination (RFE) was used to identify the most influential water parameters for classification accuracy. The results indicated that pH and total hardness were the least influential for drinking water, while SSP was the least for irrigation water. The authors acknowledge the potential use of deep learning models, which were not employed in this work. Future research could incorporate deep learning models, such as various neural network variants, as well as explore the use of unsupervised ML models as alternatives to manually calculated water quality indices. Other approaches, such as multi-criteria decision making, could also be considered for identifying influential parameters. Additionally, future work could include usage prediction models, microbial monitoring, and tracking sources of water contaminants to further enhance this research.

VI. REFERENCES

[1] B. X. Lee, F. Kjaerulf, S. Turner, L. Cohen, P. D. Donnelly, R. Muggah, R. Davis, A. Realini, B. Kieselbach, L. S. MacGregor, I. Waller, R. Gordon, M. Moloney-Kitts, G. Lee, and J. Gilligan, ‘‘Transforming our world: Implementing the 2030 agenda through sustainable development goal indicators,’’ J. Public Health Policy, vol. 37, no. S1, pp. 13–31, Sep. 2016.

[2] Integrated Approaches for Sustainable Development Goals Planning: The Case of Goal 6 on Water and Sanitation, U. ESCAP, Bangkok, Thailand,2017.

[3] WHO. Water. Protection of the Human Environment. Accessed: Jan. 24, 2022. [Online]. Available: www.afro.who.int/health-topics/water

[4] L. Ho, A. Alonso, M. A. E. Forio, M. Vanclooster, and P. L. M. Goethals, ‘‘Water research in support of the sustainable development goal 6: A case study in Belgium,’’ J. Cleaner Prod., vol. 277, Dec. 2020, Art. no. 124082.

[5] Global Nutrition Report 2016: From Promise to Impact: Ending Malnutrition by 2030, International Food Policy Research Institute, Washington, DC, USA, 2016, doi: 10.2499/9780896295841.

[6] N. Akhtar, M. I. S. Ishak, M. I. Ahmad, K. Umar, M. S. Md Yusuff, M. T. Anees, A. Qadir, and Y. K. A. Almanasir, ''Modification of the water quality index (WQI) process for simple calculation using the multicriteria decision-making (MCDM) method: A review,'' Water, vol. 13, no. 7, p. 905, Mar. 2021.

[7] World Health Organization. (1993). Guidelines for Drinking-Water Quality. World Health Organization. Accessed: Jan. 12, 2022. [Online]. Available: http://apps.who.int/iris/bitstream/handle/ 10665/44584/9789241548151-eng.pdf

[8] Standard Methods for the Examination of Water and Wastewater, Federation WE, APH Association, American Public Health Association (APHA), Washington, DC, USA, 2005.

[9] L. S. Clesceri, A. E. Greenberg, and A. D. Eaton, ''Standard methods for the examination of water and wastewater,'' Amer. Public Health Assoc. (APHA), Washington, DC, USA. Tech. Rep.21, 2005.