IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Tessellation Model For Monitoring And Assessing Water Quality For Drinking And Irrigation Purposes.

Mrs. M. Sonia*[1], Nuthala Mythri*[2], Tadi Sai Kiran*[3], Kancharla Durga Prasad*[4]

Assistant Professor, Department of Computer Science & Engineering*[1]

Scholar, Department of Information Technology *[2,3,4]

Nalla Narasimha Reddy Education Society's Group of Institutions*[1,2,3,4]

***Abstract*:** Water is a vital resource for the survival of humans, animals, and plants. However, its quality is not always suitable for drinking or other uses due to factors like industrialization, pollution, and natural occurrences. The World Health Organization has guidelines to determine the threshold levels of various parameters in water samples. The Water Quality Index (WQI) and Irrigation WQI (IWQI) are used to assess overall water quality. Collecting and measuring water samples can be challenging, but a proposed network architecture uses real-time data collection and machine learning tools to automatically determine the suitability of water samples for drinking and irrigation. The network is based on LoRa and considers the land topology. Simulations indicate a partial mesh network topology is most suitable. Since there is a lack of large datasets on drinking and irrigation water, new datasets were developed for training machine learning models. Three models (Random Forest, Logistic Regression, and Support Vector Machine) were evaluated, with Logistic Regression performing best for drinking water and Support Vector Machine for irrigation water. Recursive feature elimination was used to identify the water parameters with the greatest influence on classification accuracies.

***Keywords*:** drinking water, irrigation water

I. Introduction

The project focuses on developing a machine learning-based water quality detection system for both irrigation and drinking purposes. Recognizing the critical impact of poor water quality on crops, livestock, and human health, the system will utilize input data from various water parameters. Machine learning models, including Support Vector Machines, Naive Bayes, Logistic Regression, SGD Classifier, K-Nearest Neighbors, and Decision Trees, will be employed to analyze and predict key water quality parameters such as pH, temperature, turbidity, and dissolved oxygen. The goal is to create an accurate and efficient system capable of handling large volumes of data, identifying patterns that may be challenging for humans to discern, and providing reliable predictions for ensuring water safety. The primary objective of the project is to contribute to the safety of irrigation and drinking water. By accurately predicting water quality parameters, the system aims to detect potential issues or anomalies promptly, enabling timely interventions to prevent adverse impacts on agriculture, livestock, and human health. Successful implementation may necessitate interdisciplinary collaboration, involving experts in water quality, machine learning, and data analysis to ensure a holistic approach to system development. Ultimately, the project aspires to provide a robust and efficient solution to address critical concerns related to water quality in both agricultural and drinking water contexts.

II. Literature Survey

A water monitoring network was established in a metal-producing city in Brazil with twelve stations measuring physio-chemical water parameters. The collected data was analyzed using principal component analysis. The WaterNet project aims to collect water parameter data from city dams and proposes the use of machine learning models to automatically determine water potability or fitness for agricultural use, reducing costs and complexities associated with traditional water sample collection and analysis. This innovative approach not only streamlines the assessment process but also offers a cost-effective and efficient solution to water quality monitoring in the city. The WaterNet network addresses challenges related to sparse data transmission in water monitoring networks by emphasizing the significance of lightweight communication protocols capable of efficiently transmitting small data over long distances. This network specifically concentrates on collecting data from dams throughout the city, with the primary goal of assessing water quality for both drinking and irrigation purposes.

III. Existing System

Water monitoring involves collecting periodic samples to measure physico-chemical and microbiological metrics such as temperature, pH, and sodium levels. These measurements are sent to a base station for decision-making using lightweight communication protocols like Low Power Wide Area Network (LPWAN) technologies. While the effectiveness of simulation versus real-world testing is still debated, researchers found that NS3 simulations yielded consistent results with real-world tests using benchmark metrics such as propagation loss, coverage Packet Inter-reception (PIR), Packet Delivery Ratio (PDR), and Received Signal Strength Indicator (RSSI) level. Water parameter monitoring involves periodic sampling of physico-chemical and microbiological metrics like pH, temperature, and sodium levels. In water monitoring networks, data from measured parameters must be transmitted to a base station for decision-making. Due to sparse data transmission, lightweight communication protocols, particularly Low Power Wide Area Network (LPWAN) technologies, are favored. The efficacy of software simulations versus real-world testing in communication technologies has been debated. Researchers used NS3 for simulation and Arduino UNO C Dragino LoRa module for real-world tests, evaluating metrics like Propagation loss, coverage Packet Inter-reception (PIR), Packet Delivery Ratio (PDR), and Received Signal Strength Indicator (RSSI). They found simulation results to be consistent with real-world tests, suggesting the reliability of simulation outcomes in assessing communication technologies.

IV. Proposed System

The proposed water network aims to monitor water parameters in storage dams and treatment plants across the region. Collected data undergoes Machine Learning (ML) analysis to assess its suitability for consumption or irrigation. The project involves curating sample-sized datasets for drinking and irrigation water, which serve as training and testing sets for ML models. The data curation phase focuses on collecting suitable datasets from relevant websites. Subsequent data preprocessing involves cleaning, removing null entries, and prioritizing units and metrics. Data labeling is performed based on calculated Water Quality Index (WQI) and Irrigation Water Quality Index (IWQI) values. In the training phase, ML models process labeled data. Feature extraction identifies crucial parameters influencing model accuracy, leading to retraining for improved classification accuracy. The holistic approach integrates data collection, preprocessing, labeling, training, and feature extraction to optimize the ML models for accurate water quality assessment.

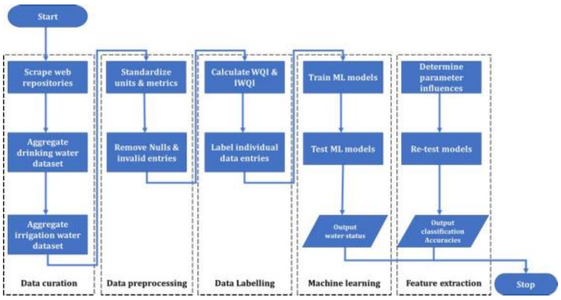

Fig 1: Block diagram of the proposed system

In the absence of large, dedicated, and openly accessible datasets for drinking and irrigation water in Africa, the researchers undertook the creation of their own datasets for machine learning (ML) research. They aggregated smaller datasets primarily from Elsevier's Data in Brief (DiB), an open-access journal focusing on research data. Using search phrases like "irrigation water," "potable water," "groundwater," and "drinking water," irrelevant articles were filtered out, resulting in 11 publications, mostly from Asia, with seven containing irrigation water data. The datasets were compiled, scraped, and organized into two CSV files, one for drinking water and another for irrigation, using Microsoft Excel. While acknowledging the ideal role of a water monitoring network as a data source for water parameters, the researchers improvised due to the lack of such a network. The primary focus at this proof-of-concept stage is on obtaining usable data for training and testing ML models, with less emphasis on the source's origin. This initiative showcases adaptability and resourcefulness in addressing data challenges for advancing ML applications in water quality assessment.

Preprocessing: The provided text outlines the methodology adopted for processing a dataset focused on drinking and irrigation water quality. With only 16% of entries pre-labelled, a Python script was developed for calculating the Water Quality Index (WQI) and Irrigation Water Quality Index (IWQI) for unlabelled data. Unlike conventional practices, equal weights were assigned to all parameters in the script to maintain a generic model devoid of bias. The labelling process involved cross-referencing WQI values with a threshold of 50, classifying values below 50 as potable (1) and above 50 as non-potable (0) for drinking water. A similar approach was taken for IWQI in the context of irrigation water. The chosen threshold of 50 was justified as a widely accepted standard in literature, denoting good or excellent water quality. However, the text underscores the need for caution, noting that while an overall WQI may indicate fitness for use, it may not capture constituents beyond the threshold, such as toxicity. Table 3 is referenced as a summary of acceptable value ranges for each parameter in the labelling script, and future subsections are teased for a more in-depth exploration of the calculation processes for WQI and IWQI.

```
1 X.head(10)
```

	id	pH	Sodium	Magnesium	Calcium	Chloride	Potassium	Carbonate	Sulphate	TDS	EC	TH	WQI	Potability
0	L1	7.3	77.5	32.6	81.4	63.6	2.0	419.7	68.2	640.0	1045.0	338.0	70.597120	0.0
1	L2	7.5	36.1	31.1	42.4	19.5	1.6	273.3	57.1	400.0	645.0	234.0	48.298251	1.0
2	L3	7.7	119.4	19.0	58.2	60.7	1.2	195.2	220.8	640.0	998.0	224.0	55.761878	0.0
3	L4	6.0	117.1	29.7	145.4	37.3	2.7	580.7	190.1	950.0	1516.0	486.0	97.028982	0.0
4	L5	7.0	45.1	23.4	49.6	23.4	1.2	248.9	74.9	410.0	658.0	220.0	46.462197	1.0
5	L6	7.0	73.4	33.0	50.4	35.2	2.3	346.5	83.0	540.0	875.0	262.0	58.783534	0.0
6	L7	7.2	121.9	29.7	67.8	47.2	1.2	195.2	310.1	740.0	1153.0	292.0	63.377606	0.0
7	L8	7.3	54.1	16.6	56.6	24.5	1.2	297.7	45.1	420.0	676.0	210.0	48.343252	1.0
8	L9	7.3	55.2	25.3	68.6	45.1	2.0	307.4	74.9	510.0	817.0	276.0	57.078475	0.0
9	L10	7.2	59.8	32.6	57.4	31.2	3.1	341.6	80.2	530.0	848.0	278.0	59.604469	0.0

Fig 2: Snippet of labelled drinking water dataset showing calculated WQI

```
1 A.head(10)
```

	id	RSC	PI	KR	MH	Na	SAR	SSP	EC	IWQI	USABLE
0	1	0.12	59.2	0.5	39.8	33.6	1.8	33.3	1045	63.379762	1.0
1	2	-0.21	58.9	0.3	54.8	25.6	1.0	25.1	645	50.337302	1.0
2	3	-1.28	72.2	1.2	35.0	53.8	3.5	53.7	998	68.144246	1.0
3	4	-0.20	55.2	0.5	25.2	34.7	2.3	34.4	1516	64.631448	1.0
4	5	-0.33	62.5	0.4	43.8	31.1	1.3	30.8	658	51.432242	1.0
5	6	0.43	66.0	0.6	52.0	38.2	2.0	37.8	875	71.445437	1.0
6	7	-2.64	63.6	0.9	42.0	47.7	3.1	47.6	1153	51.132837	1.0
7	8	0.68	69.6	0.6	32.6	36.2	1.6	35.9	676	65.043849	1.0
8	9	-0.48	58.7	0.4	37.9	30.7	1.4	30.3	817	50.248115	1.0
9	10	0.04	60.9	0.5	48.4	32.5	1.6	31.9	848	60.917361	1.0

Fig 3: Snippet of labelled irrigation water dataset showing calculated IWQI

In this section, the text discusses the utilization of hypothetical datasets curated from existing online sources, deviating from the expected data provision by WaterNet. The data curation process is detailed, highlighting the use of Algorithm 1 to calculate both the Water Quality Index (WQI) and Irrigation Water Quality Index (IWQI) for all entries. The resulting WQI values, ranging from 0 to 100, were subsequently mapped to determine the fitness of the water for drinking or irrigation purposes. Figures 2 and 3 visually present excerpts from the final labelled datasets. The penultimate column in both figures displays the calculated WQI values, while the last column indicates the assessed fitness of use for each entry. This methodology provides insight into the overall water quality and suitability for specific purposes based on the generated indices.

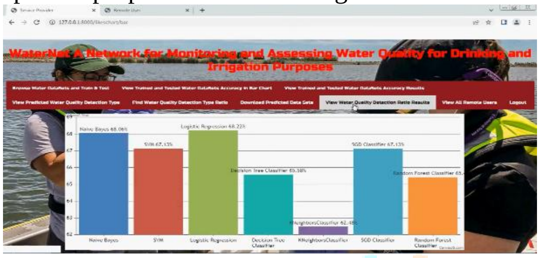

Fig 4: View Water Quality Detection Ratio Results

Figure 4 shows a graphical depiction of the result of carrying out RFE on each of the models considered, that is, RFE on LR (RFECLR), RFE on RF (RFECRF), and RFE on SVC(RFECSV). The result, though non-uniform, revealed that pH was the least influential parameter across board.

Fig 5: view accuracy results Figure 5 shows a graphical depiction of the results of recursive feature elimination (RFECLR, RFECRF, and (RFECSVC) on the irrigation water dataset. It reveals that SSP had the least influence on the classification accuracies
of the models, while RSC was the most influential feature (water parameter). SAR and Na were also relatively influential across board. EC

V. Algorithm and Technique

Algorithm 1 Calculating WQI for Drinking Water.

1. Select relevant parameters.

$(P \text{ D } [P_1; P_2; P_3 : : : P_n])$:

2. Assign weights to each parameter (w_p), $1 < p < n$
3. Calculate relative weight

$$W_p = \frac{w_p}{\sum_{p=1}^{n} w_p}$$

4. Calculate quality index

$$q_p = \frac{C_p}{S_p * 100}$$

5. Obtain

$$\text{WQI} = \sum_{p=1}^{n} W_p * q_p$$

where P = parameter selected, w_p = weight of parameter p, n = number of parameters, C_p = concentration of p, S_p = standard value for parameter p as stipulated

Algorithm 2 Calculating WQI for Irrigation Water

1. Identify prominent parameters in the sample, i.e. EC sodium, chloride, bicarbonate, SAR.
2. Determine weights for each parameter.

2a. Calculate quality measurement value

$$q_i = q_{max} - \frac{(X_{ij} - X_{inf}) * q_{iamp}}{X_{amp}}$$

2b. Calculate aggregate weight

$$w_i = \frac{\sum_{j=1}^{k} F_j * A_{ij}}{\sum_{j=1}^{k} \sum_{j=1}^{k} F_j * A_{ij}}$$

3. Obtain

$$IWQI = \sum_{i=1}^{n} q_i * w_i$$

where q_{max} = max value of q_i in its class; X_{ij} is the value of parameter i; X_{inf} is the lowest value in the class to which X_{ij} falls; q_{iamp} = class amplitude; X_{amp} = amplitude of X_{ij}'s class; w_i = parameter weight; F = autovalue of the _rst component; A_{ij} = explainability of parameter i by j; j = factor
count.

VI. Conclusion

This study addresses two primary objectives: proposing a real-time water monitoring network utilizing LoRa technology for collecting water parameter data, and applying machine learning (ML) models to assess water quality. The developed LoRa-based monitoring network, simulated in Radio Mobile, favors a partial mesh topology for optimal citywide coverage. The gathered data is intended to be aggregated on a Cloud server, where ML models—Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM)—are trained and tested. LR performs best for drinking water, boasting high accuracy and minimal false positives/negatives, while SVM excels for irrigation water. The study employs recursive feature elimination (RFE) to identify pH and total hardness as least influential for drinking water and Suspended Solids (SSP) as least influential for irrigation. Although the study

acknowledges the potential application of deep learning models, it focuses on ML models and suggests their consideration in future work. Additionally, the manual calculation of water quality indices prompts the exploration of unsupervised ML models as alternatives. Alternative approaches, such as multi-criteria decision-making, could replace RFE for identifying influential parameters. The study proposes extending the research by incorporating usage prediction models, microbial monitoring, and tracking water contaminant sources within the water network.

VII. 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, ``Modi_cation 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.
[10] M. F. Howladar, M. A. Al Numanbakth, and M. O. Faruque, ``An application of water quality index (WQI) and multivariate statistics to evaluate the water quality around Maddhapara granite mining industrial area, Dinajpur, Bangladesh,'' *Environ. Syst. Res.*, vol. 6, no. 1, pp. 1_8,Jan. 2018. 48335
[11] A. R. Finotti, R. Finkler, N. Susin, and V. E. Schneider, ``Use of water quality index as a tool for urban water resources management,'' *Int. J. Sustain. Develop. Planning*, vol. 10, no. 6, pp. 781_794, Dec. 2015.
[12] A. R. Finotti, N. Susin, R. Finkler, M. D. Silva, and V. E. Schneider, ``Development of a monitoring network of water resources in urban areasas a support for municipal environmental management,'' *WIT Trans. Ecol. Environ.*, vol. 182, pp. 133_143, May 2014.
[13] M. Chilundo, P. Kelderman, and J. H. O'keeffe, ``Design of a water quality monitoring network for the limpopo river basin in Mozambique,'' *Phys.chem. Earth, A/B/C*, vol. 33, nos. 8_13, pp. 655_665, Jan. 2008.
[14] M. Karamouz, M. Karimi, and R. Kerachian, ``Design of water quality monitoring network for river systems,'' in *Critical Transitions in Water and Environmental Resources Management*. London, U.K.: IWA, 2004, pp. 1_9.