International Journal of Technology Engineering Arts Mathematics Science
Vol. 4, No. 1, August 2024, pp. 17~24
ISSN: 2583-1224,

HydroSense: Enhancing Water Quality Monitoring using IoT and Machine Learning

Prof. Pritesh Patil[1], Anjali Mandlecha[2]
[1,2] Information Technology, AISSMS Institute of Information Technology, Maharashtra, India
Corresponding Author: Anjali Mandlecha (anjalimandlecha77@gmail.com)

Article Information

Article history:

Received Jun , 2024
Revised Jun , 2024
Accepted July , 2024

ABSTRACT

Water is necessary for everyone, including plants, animals, and humans. The hazard to living things, however, has come from the decades-long rise in water contamination. Water quality is impacted by a multitude of variables, including pollution, industrialization, and natural disasters, which alter its properties and make it less suitable for human consumption. Conventional techniques for assessing and monitoring water quality are incredibly time- and labor-intensive. Using Internet of Things sensors and Machine Learning algorithms, the research develops HydroSense, a system/model that automates water monitoring and evaluation, to get around this problem. For both irrigation and drinking, the World Health Organization has established pre-set guidelines for several of the factors found in water. Water Quality Index (WQI) and Irrigation (IWQI) are the measures used for the same. In accordance with the specifications of these metrics, HydroSense concentrates on measuring the pH, turbidity, and temperature of the water sample. The goal of the system is to automatically assess whether water is suitable for drinking or irrigation by collecting and monitoring real-time data on water characteristics. If any deviations from the water standards are found, it generates notifications on an easy-to-use online page. Users can also readily access and display data with the help of the HydroSense online portal.

KEYWORDS: pH, Temperature, Turbidity, Node MCU Wi-Fi Module, Real-time

1. INTRODUCTION (11PT)

Water is the most important asset on our planet, essential for life, agriculture, industry, and ecosystem health. With the increasing pressure on global water resources due to growth in population, climate change, and urbanization, also nowadays need for efficient and accurate water monitoring has never been greater. Poor water quality puts industrial areas at danger, harming the environment overall and resulting in financial loss. Standards for water parameters have been developed by numerous organizations, such as the WHO and BIS, and used to effectively assess the purity of water. In the past, collecting samples and sending them to a lab for testing was necessary to determine the quality of the water.

This was a lengthy operation. It is simple to acquire the values obtained from sensors from samples, monitor, and forecast the water quality from the comfort of our homes using IoT and machine learning algorithms. The sensors like pH, turbidity and temperature sensors can be used to collect the data and find out the level of danger in water using Machine Learning. Machine learning can be used for predictions using algorithms, including logistic regression, Support Vector Machine (SVM), Decision Tree (DT), K-

Nearest Neighbor (KNN), Random Forest (RF), etc. have been developed. The traditional approach to monitoring water quality entails gathering water by hand from several locations and testing it in a lab. ThingSpeak server and LED is used for the visualization purpose.

1.1. MEASUREMENT PARAMETERS OF SYSTEM

Various parameters are used to measure quality and content in water. Below are mentioned the key water parameters:

- pH value
- Turbidity
- Temperature

2. METHODOLOGY

Three sensors are used by the system: an ESP01 Wi-Fi module (NodeMCU) for data transfer, a pH, Turbidity, and Temperature sensors unit for the main processing module. Because the Arduino Mega has a lesser power consumption and a small size that makes it suitable for a critical criterion of point-of-sale technology, it is an important unit of the system designed for measuring quality of water. By employing the connection with Wi-Fi data of module ESP01 (NodeMCU) to communicate with the main server, the MCU processes and updates all the sensor data to the ThingSpeak server. The dataset is exported from ThingSpeak for prediction. Various Machine Learning models are used for prediction of the water. Below is the block diagram of the model implemented. (Figure 1)

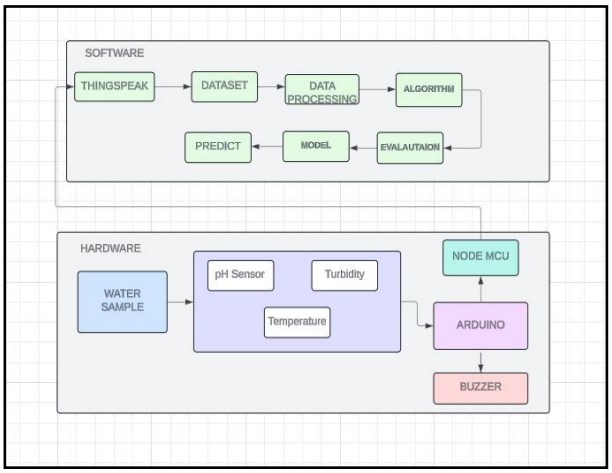

Figure 1. System's Block Diagram

The complete system/model is created in Embedded-C and the code is written in Arduino IDE. The WQM system uses sensors to collect information on variables such as pH, turbidity, water level, and temperature. By logging onto their accounts and providing a user id and password, authorized individuals can observe these data by gaining access to the ThingSpeak server. Data is collected, saved, processed, and sent instantaneously.

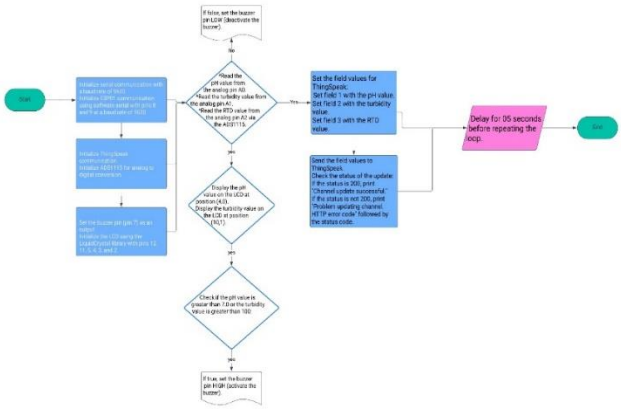

Figure 2. Complete algorithm of the model implemented

The ESP01 having feature of low-cost and is a Wi-Fi made by M/S Espino comprising of a microcontroller chip and a full TCP/IP stack of Wi-Fi chip. As it uses optimized cache capacity, the code boots straight from external flash during program implementation, increasing system rate and requiring more capacity.

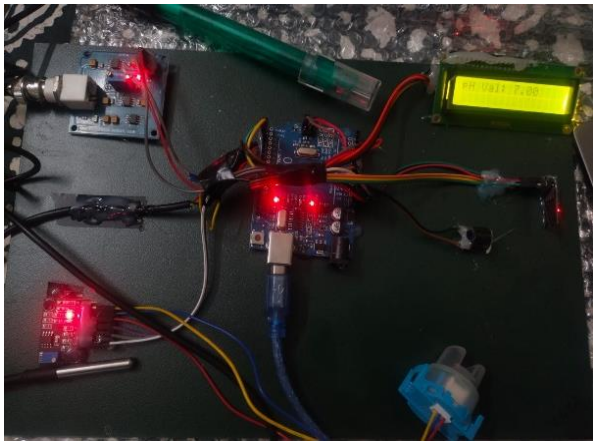

Figure 3. Hardware setup of WQM system.

During the communication, only while it is connected, the Wi-Fi module and a microcontroller need two pins (Tx/Rx), linked in the opposite direction.

4.1. ARDUINO MEGA

The Arduino Mega is powered by the 'ATmega2560 microcontroller having a robust set of features including sixteen analog inputs, a connection of USB, 4 USARTs and a 16 MHz clock generator crystal oscillator. Users can easily connect the AC/DC adapter or battery via the USB cable.

4.1.1. NodeMCU

IoT platform which is open source, integrating the Espressif systems ESP01 Wi-Fi chip-on-chip (SoC) and hardware-based ESP-12 module. With capabilities

of Wi-Fi, analog and digital pins, as well as serial communication protocols, it facilitates seamless wireless communication.

4.2. SENSORS

Variety of sensors are been used in this system are described in detail below.

4.2.1 pH SENSOR

The pH of a solution serves as a metric for its hydrogen ion concentration, showing its acidity or alkalinity level. This value is calculated by taking the negative logarithm of the hydrogen ion concentration. Typically, water has a neutral pH of around 7, with a safe pH range of drinking water between 6.5 and 8.5. A pH number less than seven denotes acidity, whereas a number more than seven denotes alkalinity. Every pH unit that increases or decreases represents a ten-fold change in hydrogen ion concentration, leading to a shift in acidity or alkalinity. pH sensors consist of measurement and reference electrodes, with the measurement electrode sensitive to hydrogen ion concentration. The electrical potential difference measured by the pH sensor is temperature dependent, necessitating the use of a temperature sensor for calibration and correction purposes.

4.2.2. TURBIDITY SENSOR

The transparency of water—which indicates the presence of suspended particles—is referred to as turbidity. Turbidity calculations gauge the amount of suspended water particles, such as clay, sand, silt, and plant detritus which affect light penetration. Elevated turbidity levels can hinder marine life reproduction and pose risks to human health The sensor has digital and analog outputs. The operating input voltage is 5V, the output voltage in analog range is 0 to 4.5V, and it able to tolerate high temperatures from 100°C to 900°C. Turbidity is typically calculated in units called nephelometric turbidity units (NTU). The sensor is located near the light source and detects the light scattered by objects suspended in water.

4.2.3. DHT-11 SENSOR

Utilizing a negative temperature coefficient (NTC) for temperature measurement, this sensor features an output of 8-bit microscope, and enables seamless transmission of temperature as well as humidity data send to microcontroller. Calibrated from the factory, it has eliminated the need for recalibration, and ensures convenient accessibility. Operating within a range of temperatures of 0°C to 50°C and humidity range of 20% to 90%, it offers an accuracy of ±1°C. Primarily employed for air temperature measurement, it ensures prolonged functionality of pH and turbidity sensors.

4.2.4 ThingSpeak SERVER

ThingSpeak functions as a comprehensive Internet of Things (IoT) platform tailored for collecting and analyzing data from a diverse array of sensors, spanning pH, turbidity, voltage, temperature, moisture, and distance sensors. It makes it easier to get data from edge node devices—like NodeMCU/ESP021—and provides extensive features for conducting past data analysis within a designed software ecosystem. Users must 1st log into their ThingSpeak accounts to access the service. The construction of channels, which include data fields as well as a status field, is the fundamental unit of ThingSpeak. Once channel is configured, MATLAB code can be used to modify, process, and interpret the data that has been collected. Users of ThingSpeak can also program alerts, such tweets, based on the data analysis.

Table 1: Data on Water Quality Index

Parameters Used	**Standard** (by WHO)
pH	6.5 – 8.5
Turbidity	<=500
Temperature (°C)	<35 °C

4.3 MODEL FOR MACHINE LEARNING

4.3.1 STANDARD VALUES

We have used the given parameters for analysis: pH, temperature, and turbidity. The appropriate values for this parameter are shown in Table 1

4.3.2 MODEL CREATION

Decision tree: DT is often used to investigate resources and purity of ground-water. From mentioned algorithm can get the relationship between different inputs and outputs and generate relationship of each according to specific standards. [13] Decision tree is the design of many learning groups. When used for demonstration purposes, it is called a connecting tree, and when used for repetition, it is called a repeating tree. Compared to other sorting techniques that integrate multiple highlights in a selection into groups, selection trees are based on selection plans or tree models at multiple levels or levels and have aggregates and joins. [16].

Random forest: The RF technique stands out as a widely embraced approach for to address regression and classification challenges. Typically, the correlation and statistical errors employed within the RF model align closely with those utilized in prior optimization methods. Nonetheless, the straightforward nature of RF for data classification renders it a favored prediction technique. [14]. RF also has advantages in terms of its capabilities and performance for generalizing standards for identifying regions possessing high-purity of drinking groundwater [13].

KNN: A non-parametric, supervised classifier, the nearest neighbor the (KNN) algorithm classifies or forecasts the grouping of individual data based on proximity. Although the KNN algorithm can be used for classification or retrieval problems, it is mainly

used for the former, based on the assumption that comparison points may be close to each other. Distance (p=2) is only valid for real vectors. Measures the straight line of the question area and other measurement points using the formula below.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(yi - x_i)^2}$$

LOGISTIC REGRESSION: A machine learning technique called logistic regression can predict the probability of an event, or observation completing a binary distribution problem. Logistic regression separates data into groups by examining the correlation between one or more variables. It is often used for predictive modeling, where the model calculates the number of events that do or do not fall into a category.

SVM: The classifier of SVM is also a popular algorithm used for predictive binary or multi-class classification as well as regression in machine learning. SVM makes the least errors in binary classification because it uses high dimensional space and uses kernel function to detect hyperplanes [15]. In the binary classification, the SVM creates the least amount of error since it employs a kernel function to find a hyperplane in a high-dimensional space [15].

Confusion matrix: This study focuses on comparison of the execution of 4 different ML classifiers to determine the best algorithm for accurate classification. The results of the product are calculated against the same data using five validity measures (accuracy, correctness, desirability, Specificity, and F1 score), where the value is one for the confusion matrix. Figure 7 depicts confusion matrices for four prediction models as examples. [15].

Precision (P): Precision is stated as the percentage of positive predictions which are by rightly identified machine learning model. This is represented as follows:

$$P = \frac{T_p}{T_p + F_p} \quad \ldots(I)$$

Recall (R): It estimates the ratio of accurately prediction quality to all appropriate models determined to be positive by the network model. Equation (II) evaluates the return value,

$$R = \frac{T_p}{T_p + F_n} \quad \ldots(II)$$

F-score: To account for both metrics, the F-score calculates the harmonic mean of precision (P) and recall (R). The F-score for the classification model is defined by equation (3):

$$F\text{-score} = 2 * \frac{(P*R)}{(P+R)} \quad \ldots(III)$$

Accuracy: This evaluates the percentage of true positive and negative class labels to the total number of class instances (Ts) in the training data. Equation (4) estimates the f-score.

$$\text{Accuracy} = \frac{T_p + T_n}{T_s} \ldots(IV)$$

5. THE PROPOSED SYSTEM ALGORITHM

In the figure 2 the algorithm used in the model/system is been shown. Firstly, the serial monitor of Arduino is launched. After that ESP Wi-Fi module along with the ThingSpeak Server is launched. The three values are read by the sensors and sensors are well connected.
The analog values are read by the Temperature Sensor. Later, the same is updated in the LED and sent to the ThingSpeak server. The necessary parameters are entered moment the ThingSpeak server receives the readings from the pH sensor after data processing has begun and displayed on the LED. Turbidity sensor is initialized and the values are displayed on LED and on the ThingSpeak server in the form of graph.

Later, the buzzer is initialized which buzzes when turbidity is high or temperature is more than 45ºC. Algorithms of Machine Learning are used to make the prediction model for the water samples collected. The data (csv file) is been exported from the ThingSpeak Server and algorithms like Decision tree, Random Forest, SVM are used for prediction purpose.

6. RESULTS OF THE MODEL

6.1. pH sample

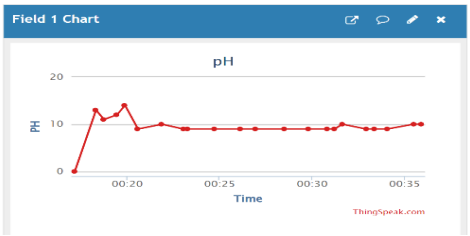

Figure 4: pH value visualization

6.2. Turbidity sample

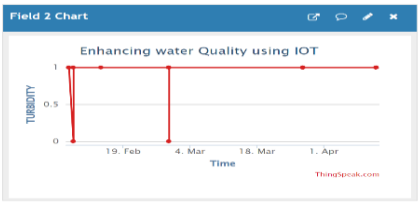

Figure 5: Turbidity value visualization

6.3. Temperature Sample

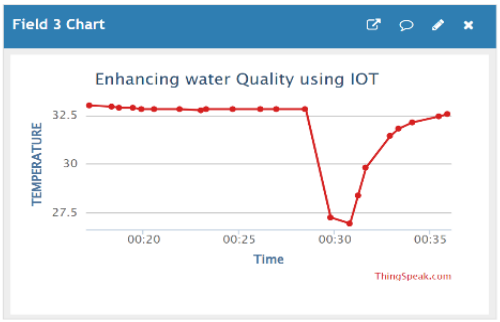

Figure 6: Temperature value visualization

6.4 pH graph

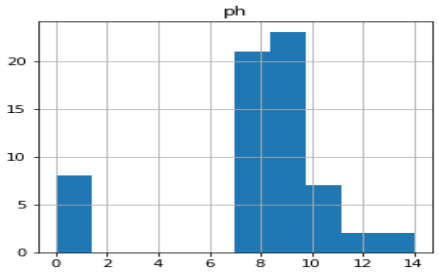

Figure 7: pH visualization

6.5. Temperature graph

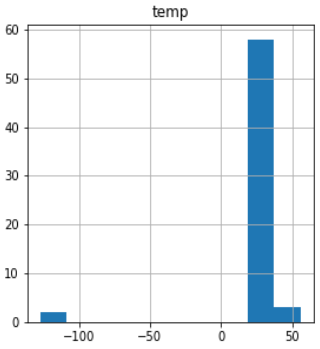

Figure 8: Temperature visualization

6.6. Turbidity graph

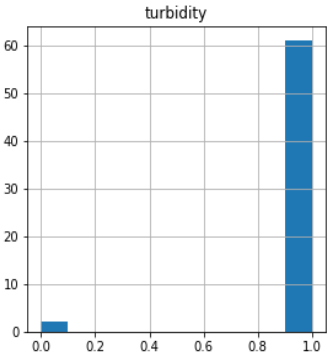

Figure 9: Turbidity visualization

Table 2: performance evaluation

Sr No.	ALGORITHM	TRAIN ACCURACY	TEST ACCURACY
1	KNN	97.72	94.73
2	LLR	100	94.732
3	DT	100	97.73
4	RF	97.73	94.73
5	SVM	94.72	94.73

Table 3: Accuracies of each algorithm

Sr No.	ALGORITHM	MSE	RMSE	MAE	R2
1	KNN	0.0526	0.2294	0.0526	-0.055
2	LR	0.054	0.324	0.052	-0.048
3	DT	0.0526	0.294	0.0226	-0.021
4	RF	0.0524	0.224	0.0626	-0.066
5	SVM	0.0526	0.229	0.0426	-0.054

***The measured values are fluctuating and can vary.

7. CONCLUSION

HydroSense is the system of combination of machine learning and Internet of Things technologies are a step forward in monitoring water purity. ML helps predict water issues early using algorithms like “Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF) and KNN.” HydroSense uses sensors for temperature, turbidity, and pH to gather important data about the water. This data is then shown on platforms like ThingSpeak and LED displays, making it easy for everyone to understand. From the results, we can conclude that Logistic Regression is best suited algorithm for the system.

8. FUTURE SCOPE

In future research directions could investigate the integration of unsupervised machine learning algorithms like clustering algorithms and anomaly detection methods could be used for

identifying subtle patterns and anomalies in water quality data. By leveraging emerging technologies like IoT grouped with cloud computing and geographic information systems (GIS), we can holistically approach the multifaceted challenges of water management in both urban and rural settings. Deep learning (DL) models like Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) could be integrated to further enhance the accuracy and efficiency of this automated monitoring system.

ADDITIONAL INFORMATION

For this paper, no further information is available.

REFERENCES

[1] Mohammad Salah Uddin Chowdury, Talha Bin Emran, Subhasish Ghosh, Abhijit Pathak, Mohd. Manjur Alam, Nurul Absar, Karl Andersson, Mohammad Shahadat Hossain, IoT Based Real-time River Water Quality Monitoring System, Procedia Computer Science, Volume 155, 2019, pp. 161-168

[2] Mosavi, Amir & Sajedi Hosseini, Farzaneh & Choubin, Bahram & Goodarzi, Massoud & Dineva, Adrienn. (2020). Groundwater Salinity Susceptibility Mapping Using Classifier Ensemble and Bayesian Machine Learning Models. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.3014908.

[3] Khan, Yafra & Chai, Soo See. (2016). Predicting and analyzing water quality using Machine Learning: A comprehensive model. 1-6. 10.1109/LISAT.2016.7494106.

[4] Ajayi, Olasupo & Bagula, Antoine & Maluleke, Hloniphani & Gaffoor, Zaheed & Jovanović, Nebojsa & Pietersen, Kevin. (2022). WaterNet: A Network for Monitoring and Assessing Water Quality for Drinking and Irrigation Purposes. IEEE Access. 10. 48318-48337. 10.1109/ACCESS.2022.3172274.

[5] Ismail, Shereen & Dawoud, Diana & Ismail, Nadhem & Marsh, Ronald & Alshami, Ali. (2022). IoT-Based Water Management Systems: Survey and Future Research Direction. IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3163742.

[6] Hao, Z. (2022). Environmental Monitoring Through IoT-Enabled Sensor Networks: A Comprehensive Approach. Computational Algorithms and Numerical Dimensions, 1(3), 116-121. doi: 10.22105/cand.2022.161802

[7] Jan, F., Min-Allah, N., & Düştegör, D. (2021). Iot based smart water quality monitoring: Recent techniques, trends and challenges for domestic applications. Water, 13(13), 1729.

[8] Bogdan R, Paliuc C, Crisan-Vida M, Nimara S, Barmayoun D. Low-Cost Internet-of-Things Water-Quality Monitoring System for Rural Areas. Sensors. 2023; 23(8):3919. https://doi.org/10.3390/s23083919 2

[9] de Camargo ET, Spanhol FA, Slongo JS, da Silva MVR, Pazinato J, de Lima Lobo AV, Coutinho FR, Pfrimer FWD, Lindino CA, Oyamada MS, et al. Low-Cost Water Quality Sensors for IoT: A Systematic Review. Sensors. 2023; 23(9):4424. https://doi.org/10.3390/s23094424

[10] Pasika, Sathish & Gandla, Sai. (2020). Smart water quality monitoring system with cost-effective using IoT. Heliyon. 6. e04096. 10.1016/j.heliyon.2020.e04096.

[11] Hong, Wong & Shamsuddin, Norazanita & Abas, Pg Emeroylariffion & Apong, Rosyzie & Masri, Zarifi & Suhaimi, Hazwani & Gödeke, Stefan & Noh, Muhammad. (2021). Water Quality Monitoring with Arduino Based Sensors. Environments. 8. 6. 10.3390/environments8010006.

[12] Md Galal Uddin, Stephen Nash, Azizur Rahman, Agnieszka I. Olbert, A novel approach for estimating and predicting uncertainty in water quality index model using machine learning approaches, Water Research, Volume 229, 2023, 119422, 10.1016/j.watres.2022.119422.

[13] Zhu, Mengyuan & Wang, Jiawei & Yang, Xiao & Zhang, Yu & Zhang, Linyu & Ren, Hongqiang & Wu, Bing & Ye, Lin. (2022). A review of the application of machine learning in water quality evaluation. Eco-Environment & Health. 1. 10.1016/j.eehl.2022.06.001.

[14] Aldrees, Ali & Javed, Muhammad Faisal & Taha, Abubakr & Mohamed, Abdeliazim & Jasiński, Michał & Gono, Miroslava. (2023). Evolutionary and Ensemble Machine Learning Predictive Models for Evaluation of Water Quality. Journal of Hydrology: Regional Studies. 46. 10.1016/j.ejrh.2023.101331.

[15] Uddin, Md Galal & Nash, Stephen & Rahman, Azizur & Olbert, Agnieszka. (2022). Performance Analysis of the Water Quality Index Model for Predicting Water State Using Machine Learning Techniques. Process Safety and Environmental Protection. 10.1016/j.psep.2022.11.073.

[16] Nishant Rawat, Mangani Daudi Kazembe, Pradeep Kumar Mishra (2022). Water Quality Prediction using Machine Learning. ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538.